

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: September 18, 2004, 09:46:06 ; Search time 1142.55 Seconds  
(without alignments)  
9066.559 Million cell updates/sec  
Title: US-10-614-282-1  
Perfect score: 239  
Sequence: 1 atcagtcaagcttggtgaag.....tgtggtccggctggacaat 239

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.scs.\*  
28: em.un.\*  
29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.inv.\*  
32: em.htg.other.\*  
33: em.htg.mus.\*  
34: em.htg.pln.\*  
35: em.htg.rod.\*  
36: em.htg.mam.\*  
37: em.htg.vrt.\*  
38: em.sy.\*  
39: em.htgo.hum.\*  
40: em.htgo.mus.\*  
41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	239	100.0	1846	3	DMLABG1
2	239	100.0	2791	3	AY095089 Drosophila
3	239	100.0	63442	2	AC020280 Drosophila
C 4	239	100.0	85757	3	AC001652 Drosophila
5	239	100.0	110000	3	AC001572_3
C 6	239	100.0	170801	3	AC095014 Drosophila
C 7	239	100.0	175648	3	AC095013 Drosophila
C 8	239	100.0	298020	3	AE003674 Drosophila
9	36.6	15.3	237630	2	AC106665 Rattus no
10	36.6	15.3	239015	2	AC131194 Rattus no
C 11	36.6	15.3	282019	2	AC096362 Rattus no
C 12	34.8	14.6	20938	1	AE008718 Salmonell
C 13	34	14.2	178993	2	AC099723 Mus muscu
14	33.4	14.0	193636	2	AC113984 Mus muscu
15	33.4	14.0	229102	2	AC119849 Mus muscu
C 16	33.2	13.9	172214	2	BX294181 Danio rer
17	33.2	13.9	240050	1	AL627267 Salmonell
C 18	33.2	13.9	300029	1	AE016842 Salmonell
C 19	32.8	13.7	727	6	AX416024 Sequence
C 20	32.8	13.7	11621	1	AE003903 Xylella f
C 21	32.8	13.7	151733	2	AC136249 Rattus no
C 22	32.8	13.7	211426	2	AC119522 Rattus no
C 23	32.8	13.7	216886	2	AC127846 Rattus no
C 24	32.8	13.7	232601	2	AC112281 Rattus no
C 25	32.8	13.7	232601	2	AC112281 Rattus no
C 26	32.8	13.7	233002	2	AC120490 Rattus no
27	32.8	13.7	234803	2	AC094061 Rattus no
C 28	32.8	13.7	236723	2	AC106621 Rattus no
C 29	32.8	13.7	244293	2	AC133774 Rattus no
C 30	32.6	13.6	87077	5	BX571732 Zebrafish
C 31	32.6	13.6	115859	8	CNS08CE0 Oryza sat
C 32	32.6	13.6	129179	2	AC117205 Mus muscu
33	32.6	13.6	134334	10	AL163512 Mouse DNA
34	32.6	13.6	144486	10	AL626776 Mouse DNA
C 35	32.6	13.6	148716	5	AL935049 Zebrafish
C 36	32.6	13.6	165860	10	AL831716 Mouse DNA
37	32.6	13.6	172465	2	BX324192 Mus muscu
C 38	32.6	13.6	193230	2	AC138639 Mus muscu
C 39	32.6	13.6	231548	2	AC146669 Otollemur
C 40	32.6	13.6	246647	5	AL928716 Zebrafish
C 41	32.4	13.6	100919	9	AL359378 Human DNA
C 42	32.4	13.6	169226	2	AC107725 Mus muscu
C 43	32.4	13.6	248818	2	AC112853 Rattus no
C 44	32.4	13.6	256414	2	AC132775 Rattus no
C 45	32.4	13.6	258710	2	AC135823 Rattus no

# ALIGNMENTS

RESULT 1  
DMLABG1  
LOCUS DMLABG1 1846 bp DNA linear INV 06-OCT-1999  
DEFINITION Drosophila melanogaster lab gene for labial protein, exon 1 and  
joined CDS.  
ACCESSION X13104 X12834  
VERSION X13104.1 GI:8171  
KEYWORDS alternative splicing; antennapedia complex; homeobox; lab gene.  
SOURCE Drosophila melanogaster (fruit fly)  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 1846)  
AUTHORS Mlodzik,M., Fjose,A. and Gehring,W.J.

TITLE Molecular structure and spatial expression of a homeobox gene from the labial region of the Antennapedia-complex  
 JOURNAL EMBO J. 7 (8), 2569-2578 (1988)  
 MEDLINE 89052673  
 PUBMED 2461299  
 REFERENCE 2 (bases 1 to 1846)  
 AUTHORS Doelz, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (12-AUG-1998) Doelz R., Biocomputing Biozentrum der Universitat, Klingelbergstrasse 70, CH-4056 Basel, Switzerland  
 COMMENT see M13570 for overlapping sequence.  
 FEATURES  
 Location/Qualifiers  
 1..1846  
 /organism="Drosophila melanogaster"  
 /mol\_type="genomic DNA"  
 /strain="Oregon-R"  
 /db\_xref="taxon:7227"  
 /chromosomes="84A"  
 join(392..1846,X13105.2:41..474,X13106.1:1..258)  
 /gene="lab"  
 392..1846  
 /gene="lab"  
 /number=1  
 join(631..1846,X13105.2:41..474,X13106.1:1..258)  
 /gene="lab"  
 /codon\_start=1  
 /label=lab\_CDS1  
 /product="labial protein"  
 /protein\_id="CAB57786.1"  
 /db\_xref="GI:6015508"  
 /db\_xref="GOA:P10105"  
 /db\_xref="SWISS-PROT:P10105"  
 /translation="MMDVSMYGNPHHHHPHANAYDGYSTTTASAAANASSYFAPQOH QPHLQOQQOQHHLQPOOHLTYNGYESSPGNYPOQQAOLTPPTSSHQVVOHQH QOQOQQOQLPHSLFSPAAEYGITTTTNGPKPLHPSHPADSYESDSVHS YVATAVAVAPSSPITAAANASATNTQOQOAAIISSENGMMYNLDCMYPTA QACAPVHGAGIIEKYAAVLHASVAPGVLEDDPMMQOATQSQMWHQHLAGSYA LDAMSLGMHWHHLPHGLHNLANNPHQOQPQVQOQQOQPHQPPHQPSAAH QQHNSVSPNGMRQQRGGVSPSSSTSSASNGAHPASTQSKSPHSSSIPTY KMWJLRNKPQAPSYLPAPKLPASGIAASHDYQMNGLDMCRGGGGSDVGSVPV GYGGNGPIGVLSVQNSLIWANAASAAHNGMVGSLGSLSCSLSSNTNN SGRNTFTKQLTEKEFHFNRYLTRARRIETANTLQNETQVTFQNRMRKQKRV KGLIPADILTLQHSITSVISEKPPQOQPPPELQLKSGSLGNGELATGAFSTPTTA MTLTAPTSKQS"  
 join(631..1846,X13105.2:59..474,X13106.1:1..258)  
 /gene="lab"  
 /codon\_start=1  
 /label=lab\_CDSII  
 /product="labial protein"  
 /protein\_id="CAB57787.1"  
 /db\_xref="GI:6015509"  
 /db\_xref="GOA:P10105"  
 /db\_xref="SWISS-PROT:P10105"  
 /translation="MMDVSMYGNPHHHHPHANAYDGYSTTTASAAANASSYFAPQOH QPHLQOQQOQHHLQPOOHLTYNGYESSPGNYPOQQAOLTPPTSSHQVVOHQH QOQOQQOQLPHSLFSPAAEYGITTTTNGPKPLHPSHPADSYESDSVHS YVATAVAVAPSSPITAAANASATNTQOQOAAIISSENGMMYNLDCMYPTA QACAPVHGAGIIEKYAAVLHASVAPGVLEDDPMMQOATQSQMWHQHLAGSYA LDAMSLGMHWHHLPHGLHNLANNPHQOQPQVQOQQOQPHQPPHQPSAAH QQHNSVSPNGMRQQRGGVSPSSSTSSASNGAHPASTQSKSPHSSSIPTY KMWJLRNKPQAPKLPASGIAASHDYQMNGLDMCRGGGGSDVGSVPV GYGGNGPIGVLSVQNSLIWANAASAAHNGMVGSLGSLSCSLSSNTNN SGRNTFTKQLTEKEFHFNRYLTRARRIETANTLQNETQVTFQNRMRKQKRV KGLIPADILTLQHSITSVISEKPPQOQPPPELQLKSGSLGNGELATGAFSTPTTA MTLTAPTSKQS"  
 join(631..1846,X13105.2:59..474,X13106.1:1..258)  
 /gene="lab"  
 /codon\_start=1  
 /label=lab\_CDSII  
 /product="labial protein"  
 /protein\_id="CAB57787.1"  
 /db\_xref="GI:6015509"  
 /db\_xref="GOA:P10105"  
 /db\_xref="SWISS-PROT:P10105"  
 /translation="MMDVSMYGNPHHHHPHANAYDGYSTTTASAAANASSYFAPQOH QPHLQOQQOQHHLQPOOHLTYNGYESSPGNYPOQQAOLTPPTSSHQVVOHQH QOQOQQOQLPHSLFSPAAEYGITTTTNGPKPLHPSHPADSYESDSVHS YVATAVAVAPSSPITAAANASATNTQOQOAAIISSENGMMYNLDCMYPTA QACAPVHGAGIIEKYAAVLHASVAPGVLEDDPMMQOATQSQMWHQHLAGSYA LDAMSLGMHWHHLPHGLHNLANNPHQOQPQVQOQQOQPHQPPHQPSAAH QQHNSVSPNGMRQQRGGVSPSSSTSSASNGAHPASTQSKSPHSSSIPTY KMWJLRNKPQAPKLPASGIAASHDYQMNGLDMCRGGGGSDVGSVPV GYGGNGPIGVLSVQNSLIWANAASAAHNGMVGSLGSLSCSLSSNTNN SGRNTFTKQLTEKEFHFNRYLTRARRIETANTLQNETQVTFQNRMRKQKRV TNGKLTLEKEFHFNRYLTRARRIETANTLQNETQVTFQNRMRKQKRVKEGLIP ADILTLQHSITSVISEKPPQOQPPPELQLKSGSLGNGELATGAFSTPTTA TSKQS"  
 ORIGIN  
 Query Match 100.0%; Score 239; DB 3; Length 1846;  
 Best Local Similarity 100.0%; Pred. No. 2,5e-66;  
 Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 1 ATCAGTCACGACTTGGTAAAGCGCGAGCAGCAGCTGCTGCTGCTATCGCCAAACGGGAG 60  
 |||||  
 1 ATCAGTCACGACTTGGTAAAGCGCGAGCAGCAGCTGCTGCTGCTATCGCCAAACGGGAG 60  
 |||||

Db 392 ATCAGTCACGACTTGGTAAAGCGCGAGCAGCAGCTGCTGCTGCTATCGCCAAACGGGAG 451  
 QY 61 TCGTGTGTTTTTCGGTTCGATACAGATAAAACCCACGTCGATAGCCTTCGACCGTCGCGTAA 120  
 |||||  
 Db 452 TCGTGTGTTTTTCGGTTCGATACAGATAAAACCCACGTCGATAGCCTTCGACCGTCGCGTAA 511  
 |||||  
 QY 121 TATCTTGAAGAACCAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGTT 180  
 |||||  
 Db 512 TATCTTGAAGAACCAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGTT 571  
 |||||  
 QY 181 ACTTAGTGATACACCGGTTTATATCGAGTGGCGAGAAAGTGTGGTTCGGCTGGACAAT 239  
 |||||  
 Db 572 ACTTAGTGATACACCGGTTTATATCGAGTGGCGAGAAAGTGTGGTTCGGCTGGACAAT 630  
 |||||  
 RESULT 2  
 LOCUS AY095089 2791 bp mRNA linear INV 15-APR-2002  
 DEFINITION Drosophila melanogaster RE63854 full insert cDNA.  
 ACCESSION AY095089  
 VERSION AY095089.1 GI:20152114  
 KEYWORDS FLI CDNA.  
 SOURCE Drosophila melanogaster (fruit fly)  
 ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 2791)  
 Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Carlson,J., Champe,M., Chavez,C., Dorsett,V., Dresnek,D., Farfan,D., Frise,E., George,R., Gonzalez,M., Guarin,H., Kronmiller,B., Li,P., Liao,G., Miranda,A., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S., Patel,S., Phouanenavong,S., Wan,K., Yu,C., Lewis,S.E., Rubin,G.M. and Celniker,S.  
 Direct Submission  
 Submitted (09-APR-2002) Berkeley Drosophila Genome Project, Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, USA  
 Sequence submitted by:  
 Berkeley Drosophila Genome Project  
 Lawrence Berkeley National Laboratory  
 Berkeley, CA 94720  
 This clone was sequenced as part of a high-throughput process to sequence clones from Drosophila Gene Collection 1 (Rubin et al., Science 2000). The sequence has been subjected to integrity checks for sequence accuracy, presence of a polyA tail and contiguity within 100 kb in the genome. Thus we believe the sequence to reflect accurately this particular cDNA clone. However, there are artifacts associated with the generation of cDNA clones that may have not been detected in our initial analyses such as internal priming, priming from contaminating genomic DNA, retained introns due to reverse transcription of unspliced precursor RNAs, and reverse transcriptase errors that result in single base changes. For further information about this sequence, including its location and relationship to other sequences, please visit our Web site (<http://fruitfly.berkeley.edu>) or send email to [cdna@fruitfly.berkeley.edu](mailto:cdna@fruitfly.berkeley.edu).  
 Location/Qualifiers  
 1..2791  
 /organism="Drosophila melanogaster"  
 /mol\_type="mRNA"  
 /strain="y; cn bw sp"  
 /db\_xref="taxon:7227"  
 /map="84A2-84A3"  
 1..2791  
 /gene="lab"  
 /note="alignment with genomic scaffold AE003674"  
 /db\_xref="FLYBASE:FBgn0002522"  
 245..2134  
 /gene="lab"  
 /note="Longest ORF"  
 /codon\_start=1  
 /product="RE63854p"  
 FEATURES  
 source  
 1..2791  
 /organism="Drosophila melanogaster"  
 /mol\_type="mRNA"  
 /strain="y; cn bw sp"  
 /db\_xref="taxon:7227"  
 /map="84A2-84A3"  
 1..2791  
 /gene="lab"  
 /note="alignment with genomic scaffold AE003674"  
 /db\_xref="FLYBASE:FBgn0002522"  
 245..2134  
 /gene="lab"  
 /note="Longest ORF"  
 /codon\_start=1  
 /product="RE63854p"  
 gene  
 1..2791  
 /gene="lab"  
 /note="alignment with genomic scaffold AE003674"  
 /db\_xref="FLYBASE:FBgn0002522"  
 245..2134  
 /gene="lab"  
 /note="Longest ORF"  
 /codon\_start=1  
 /product="RE63854p"  
 CDS  
 245..2134  
 /gene="lab"  
 /note="Longest ORF"  
 /codon\_start=1  
 /product="RE63854p"

```

/protein_id="BAM11417.1"
/db_xref="GI:20152115"
/db_xref="FLYBASE:FBgn0002522"
/translation="MMDSVSNMGPHHHPHANAYDGYSTTTASAAASGYFAPOOH
QQHQLQOQOQLVPHSHLSPSAEYGIITSTTGNPGTPLPHSPHSPADSVYSDVSH
YVATAAVATVAPPSSNPITAAASATNTQOQOQOQAAIISSENGMYATNLDCWYPTA
QAPVHYAGQIQBEKYAUVLHASYAFGVLEDDQPMQOQATQSMWHQHQLAGSVA
LDAMDGLGMAHMHGUPHGLHGLANPHQOQFOVQOQOQOQOQOQOQOQOQOQOQOQ
QOQOQNSVSPNGNRRQGGVISPSTSSSSASNGAHPASTQSPKSPHSSSIPTY
KMWQLKRNVPKPAKLPASGIASHDYQMGQDMCRGGGGGGGGVGVGVGGNG
SPGIGVLVSQNSILMANSAAGSAHPNGMGVGLSGSLSSCNTNTNSVGRITNF
TNKOLTELEKEFFENRYLTRARRTEIANTLQNETOVKMFQNRMRKQKVKKEGLIP
ADILQTSVTSVSEKFPQOQOQPPPELQLKSGSDLGNELATGAPSTPTTAMLTAP
TSKQS"

ORIGIN
Query Match          100.0%; Score 239; DB 3; Length 2791;
Best Local Similarity 100.0%; Pred. No. 2.7e-66;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATCAGTCACGACTTGGTAAGCGCGCAGCAGCAGCTGCTGCTGTCATCGCCACGGGAG 60
|
|
|
Db 6 ATCAGTCACGACTTGGTAAGCGCGCAGCAGCAGCTGCTGCTGTCATCGCCACGGGAG 65
|
|
|
Qy 61 TCCTGTTTTTCGGTTCGATACAGATAAACCACCGTCGATAGCCCTCGACGTCGCGTAA 120
|
|
|
Db 66 TCCTGTTTTTCGGTTCGATACAGATAAACCACCGTCGATAGCCCTCGACGTCGCGTAA 125
|
|
|
Qy 121 TATCTTAGAAGCAACAGCTAAAGAACTATTTCAGAACTGTGTGCGCAAGTGAAGGTT 180
|
|
|
Db 126 TATCTTAGAAGCAACAGCTAAAGAACTATTTCAGAACTGTGTGCGCAAGTGAAGGTT 185
|
|
|
Qy 181 AGTTAGTGATACACCGGTTATATCGGAGTGGCGAGAAAGTGTGTCGCGTGGACAAAT 239
|
|
|
Db 186 AGTTAGTGATACACCGGTTATATCGGAGTGGCGAGAAAGTGTGTCGCGTGGACAAAT 244
|
|
|

RESULT 3
AC020280 63442 bp DNA linear HTG 03-JAN-2000
LOCUS Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***
DEFINITION Drosophila melanogaster (fruit fly)
ACCESSION AC020280
VERSION AC020280.1 GI:5664617
KEYWORDS HTG; HGVS_PHASE2.
SOURCE Drosophila melanogaster
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 63442)
Adams,M. and Venter,J.C.
Direct Submission
Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
This sequence was identified as CDM:10212650 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
Location/Qualifiers
1..63442
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"

ORIGIN
Query Match          100.0%; Score 239; DB 2; Length 63442;
Best Local Similarity 100.0%; Pred. No. 4.6e-66;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATCAGTCACGACTTGGTAAGCGCGCAGCAGCAGCTGCTGCTGTCATCGCCACGGGAG 60
|
|
|

/protein_id="BAM11417.1"
/db_xref="GI:20152115"
/db_xref="FLYBASE:FBgn0002522"
/translation="MMDSVSNMGPHHHPHANAYDGYSTTTASAAASGYFAPOOH
QQHQLQOQOQLVPHSHLSPSAEYGIITSTTGNPGTPLPHSPHSPADSVYSDVSH
YVATAAVATVAPPSSNPITAAASATNTQOQOQOQAAIISSENGMYATNLDCWYPTA
QAPVHYAGQIQBEKYAUVLHASYAFGVLEDDQPMQOQATQSMWHQHQLAGSVA
LDAMDGLGMAHMHGUPHGLHGLANPHQOQFOVQOQOQOQOQOQOQOQOQOQOQOQ
QOQOQNSVSPNGNRRQGGVISPSTSSSSASNGAHPASTQSPKSPHSSSIPTY
KMWQLKRNVPKPAKLPASGIASHDYQMGQDMCRGGGGGGGGVGVGVGGNG
SPGIGVLVSQNSILMANSAAGSAHPNGMGVGLSGSLSSCNTNTNSVGRITNF
TNKOLTELEKEFFENRYLTRARRTEIANTLQNETOVKMFQNRMRKQKVKKEGLIP
ADILQTSVTSVSEKFPQOQOQPPPELQLKSGSDLGNELATGAPSTPTTAMLTAP
TSKQS"

ORIGIN
Query Match          100.0%; Score 239; DB 3; Length 2791;
Best Local Similarity 100.0%; Pred. No. 2.7e-66;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATCAGTCACGACTTGGTAAGCGCGCAGCAGCAGCTGCTGCTGTCATCGCCACGGGAG 60
|
|
|
Db 6 ATCAGTCACGACTTGGTAAGCGCGCAGCAGCAGCTGCTGCTGTCATCGCCACGGGAG 65
|
|
|
Qy 61 TCCTGTTTTTCGGTTCGATACAGATAAACCACCGTCGATAGCCCTCGACGTCGCGTAA 120
|
|
|
Db 66 TCCTGTTTTTCGGTTCGATACAGATAAACCACCGTCGATAGCCCTCGACGTCGCGTAA 125
|
|
|
Qy 121 TATCTTAGAAGCAACAGCTAAAGAACTATTTCAGAACTGTGTGCGCAAGTGAAGGTT 180
|
|
|
Db 126 TATCTTAGAAGCAACAGCTAAAGAACTATTTCAGAACTGTGTGCGCAAGTGAAGGTT 185
|
|
|
Qy 181 AGTTAGTGATACACCGGTTATATCGGAGTGGCGAGAAAGTGTGTCGCGTGGACAAAT 239
|
|
|
Db 186 AGTTAGTGATACACCGGTTATATCGGAGTGGCGAGAAAGTGTGTCGCGTGGACAAAT 244
|
|
|

RESULT 4
AC001652/c 85757 bp DNA linear INV 17-JUL-1998
LOCUS Drosophila melanogaster DNA sequence (Pl DS00004 (D12)), complete
DEFINITION Drosophila melanogaster (fruit fly)
ACCESSION AC001652
VERSION AC001652.1 GI:2341048
KEYWORDS HTG.
SOURCE Drosophila melanogaster
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 85757)
Celisniker,S.E., George,R.A., Galle,R.F., Hoskins,R.A.,
Svirskas,R.R., Harris,N.L., Agbayani,A., Arcaina,T.T., Baxter,E.,
Blazej,R.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E.,
Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,S.H., Lee,B., Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Punch,E., Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R.,
Yee,A., Zhang,R., Zieran,L.L. and Kimmel,B.
Sequencing of antennapedia complex, homeotic genes
Unpublished (1997)
2 (bases 1 to 85757)
Martin,C.H., Arcaina,T.T., Bondoc,M.M., Chiang,A., Critz,P.A.,
Davis,C.A., Doyle,C.M., Ericsson,C.L., Farfan,D.E., Gunning,K.M.,
Houston,K.A., Jaklevic,M.A., Kadner,K.E., Kim,K., Kim,S.F.,
Ko,C.L., Lewis,K.D., Li,M., Lindquist,K.J., Lomotan,M.A.,
Lustre,V.M., Machrus,M.U., Mayeda,C.A., Miguel,T.M., Miller,C.A.,
Mok,M.S., Pacleb,J.M., Patel,S.G., Santos,R.F., Subramanian,S.,
Wan,K.H., Whitelaw,K.R., Yee,A., Yeh,R.T., Yu,C. and Palazzolo,M.J.
Direct Submission
Submitted (22-APR-1997)
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://fruitfly.berkeley.edu/sequence/) or send
email to drosophila@hgsc.lbl.gov.
Library location: 4-1
Location/Qualifiers
1..85757
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/chromosome="3R"
/map="84A1-84A2"
/clone="Pl DS00004 (D12)"
/note="This sequence has not changed since its original
submission on 04/22/1997. It was resubmitted in order to
include all secondary accession numbers for the subclones"

```

belonging to this clone."

ORIGIN

Query Match 100.0%; Score 239; DB 3; Length 85757;  
 Best Local Similarity 100.0%; Pred. No. 4.9e-66;  
 Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCAGTCACGACTTGTTAAAGCGGCGAGCAGCAGCTGCTGCTATCCCAACGGGAG 60  
 Db 51586 ATCAGTCACGACTTGTTAAAGCGGCGAGCAGCAGCTGCTGCTATCCCAACGGGAG 51527

QY 61 TCGTGTTTTTCGGTTCGATACAGATAAAACCCAGTCGATAGCCCTCGACCGTCGGTAA 120  
 Db 51526 TCGTGTTTTTCGGTTCGATACAGATAAAACCCAGTCGATAGCCCTCGACCGTCGGTAA 51467

QY 121 TATTCCTTAGAAAGCAACAGCTAAAGAACTATTTCCTCAAGAACTGTGTGGCAAGTCAAGGGT 180  
 Db 51466 TATTCCTTAGAAAGCAACAGCTAAAGAACTATTTCCTCAAGAACTGTGTGGCAAGTCAAGGGT 51407

QY 181 AGTTAGTGATACACCGGTTATTCGGAGTCGGAGGAGAAAGTGTGTTCCGGCTGGACAAT 239  
 Db 51406 AGTTAGTGATACACCGGTTATTCGGAGTCGGAGGAGAAAGTGTGTTCCGGCTGGACAAT 51348

RESULT 5  
 AE001572.3  
 WPCOMMENT

Sequence split into 5 fragments LOCUS AE001572 Accession AE001572

Fragment Name	Begin	End
AE001572.0	1	110000
AE001572.1	100001	210000
AE001572.2	200001	310000
AE001572.3	300001	410000
AE001572.4	400001	423825

Continuation (4 of 5) of AE001572 from base 300001 (AE001572 Drosophila melanogaster Amb

Query Match 100.0%; Score 239; DB 3; Length 110000;  
 Best Local Similarity 100.0%; Pred. No. 5.1e-66;  
 Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCAGTCACGACTTGTTAAAGCGGCGAGCAGCAGCTGCTGCTATCCCAACGGGAG 60  
 Db 78240 ATCAGTCACGACTTGTTAAAGCGGCGAGCAGCAGCTGCTGCTATCCCAACGGGAG 78299

QY 61 TCGTGTTTTTCGGTTCGATACAGATAAAACCCAGTCGATAGCCCTCGACCGTCGGTAA 120  
 Db 78300 TCGTGTTTTTCGGTTCGATACAGATAAAACCCAGTCGATAGCCCTCGACCGTCGGTAA 78359

QY 121 TATTCCTTAGAAAGCAACAGCTAAAGAACTATTTCCTCAAGAACTGTGTGGCAAGTCAAGGGT 180  
 Db 78360 TATTCCTTAGAAAGCAACAGCTAAAGAACTATTTCCTCAAGAACTGTGTGGCAAGTCAAGGGT 78419

QY 181 AGTTAGTGATACACCGGTTATTCGGAGTCGGAGGAGAAAGTGTGTTCCGGCTGGACAAT 239  
 Db 78420 AGTTAGTGATACACCGGTTATTCGGAGTCGGAGGAGAAAGTGTGTTCCGGCTGGACAAT 78478

RESULT 6  
 AC095014/c

LOCUS AC095014  
 DEFINITION Drosophila melanogaster, chromosome 3R, region 83F-84A, BAC clone  
 AC095014, complete sequence.  
 AC095014.1 GI:15624856  
 HTG.  
 Drosophila melanogaster (fruit fly)  
 Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 170801)  
 REFERENCE  
 AUTHORS

181 AGTTAGTGATACACCGGTTATTCGGAGTCGGAGGAGAAAGTGTGTTCCGGCTGGACAAT 239  
 Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,  
 Celisner,S.B., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,  
 Holt,R.A., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,  
 Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,  
 Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dreenek,D., Farfan,D.,  
 Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,  
 Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,  
 Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,  
 McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,  
 Pacleeb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B., Scheeler,F.,  
 Shouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,  
 Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,  
 Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.

Submitted (15-SEP-2001) Berkeley Drosophila Genome Project, MS  
 64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,  
 Berkeley, CA 94720, US  
 Sequence submitted by:  
 Berkeley Drosophila Genome Project  
 Lawrence Berkeley National Laboratory, MS 64-121  
 Berkeley, CA 94720  
 This sequence was assembled using end sequences from a whole genome  
 shotgun and from subclones of this BAC and its neighboring clones.  
 For further information about this sequence, including its location  
 and relationship to other sequences, please visit our sequence  
 archive Web site (<http://www.fruitfly.org/sequence/>) or send email  
 to [bdgp@fruitfly.berkeley.edu](mailto:bdgp@fruitfly.berkeley.edu).

FEATURES  
 source  
 1..170801  
 /organism="Drosophila melanogaster"  
 /mol\_type="genomic DNA"  
 /strain="y; cn bw sp"  
 /db\_xref="taxon:7227"  
 /chromosome="3R"  
 /map="83F-84A"  
 /clone="BACR19D23 (D1349)"  
 /clone\_lib="RPC1-98 (Roswell Park Cancer Institute  
 Drosophila melanogaster BAC library, partial EcoRI in  
 pBACe3.6)"

ORIGIN

Query Match 100.0%; Score 239; DB 3; Length 170801;  
 Best Local Similarity 100.0%; Pred. No. 5.5e-66;  
 Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCAGTCACGACTTGTTAAAGCGGCGAGCAGCAGCTGCTGCTATCCCAACGGGAG 60  
 Db 12530 ATCAGTCACGACTTGTTAAAGCGGCGAGCAGCAGCTGCTGCTATCCCAACGGGAG 12471

QY 61 TCGTGTTTTTCGGTTCGATACAGATAAAACCCAGTCGATAGCCCTCGACCGTCGGTAA 120  
 Db 12470 TCGTGTTTTTCGGTTCGATACAGATAAAACCCAGTCGATAGCCCTCGACCGTCGGTAA 12411

QY 121 TATTCCTTAGAAAGCAACAGCTAAAGAACTATTTCCTCAAGAACTGTGTGGCAAGTCAAGGGT 180  
 Db 12410 TATTCCTTAGAAAGCAACAGCTAAAGAACTATTTCCTCAAGAACTGTGTGGCAAGTCAAGGGT 12351

QY 181 AGTTAGTGATACACCGGTTATTCGGAGTCGGAGGAGAAAGTGTGTTCCGGCTGGACAAT 239  
 Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,  
 Celisner,S.B., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,  
 Holt,R.A., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,  
 Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,  
 Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dreenek,D., Farfan,D.,  
 Ferriera,S., Frise,E., Galle,R.F., Garg,N.S., George,R.A.,  
 Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,  
 Ibegwam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,  
 McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,  
 Pacleeb,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B., Scheeler,F.,  
 Shouanenavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,  
 Stapleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,  
 Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.

Submitted (15-SEP-2001) Berkeley Drosophila Genome Project, MS  
 64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,  
 Berkeley, CA 94720, US  
 Sequence submitted by:  
 Berkeley Drosophila Genome Project  
 Lawrence Berkeley National Laboratory, MS 64-121  
 Berkeley, CA 94720  
 This sequence was assembled using end sequences from a whole genome  
 shotgun and from subclones of this BAC and its neighboring clones.  
 For further information about this sequence, including its location  
 and relationship to other sequences, please visit our sequence  
 archive Web site (<http://www.fruitfly.org/sequence/>) or send email  
 to [bdgp@fruitfly.berkeley.edu](mailto:bdgp@fruitfly.berkeley.edu).

FEATURES  
 source  
 1..170801  
 /organism="Drosophila melanogaster"  
 /mol\_type="genomic DNA"  
 /strain="y; cn bw sp"  
 /db\_xref="taxon:7227"  
 /chromosome="3R"  
 /map="83F-84A"  
 /clone="BACR19D23 (D1349)"  
 /clone\_lib="RPC1-98 (Roswell Park Cancer Institute  
 Drosophila melanogaster BAC library, partial EcoRI in  
 pBACe3.6)"

ORIGIN

Query Match 100.0%; Score 239; DB 3; Length 170801;  
 Best Local Similarity 100.0%; Pred. No. 5.5e-66;  
 Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCAGTCACGACTTGTTAAAGCGGCGAGCAGCAGCTGCTGCTATCCCAACGGGAG 60  
 Db 12530 ATCAGTCACGACTTGTTAAAGCGGCGAGCAGCAGCTGCTGCTATCCCAACGGGAG 12471

QY 61 TCGTGTTTTTCGGTTCGATACAGATAAAACCCAGTCGATAGCCCTCGACCGTCGGTAA 120  
 Db 12470 TCGTGTTTTTCGGTTCGATACAGATAAAACCCAGTCGATAGCCCTCGACCGTCGGTAA 12411

QY 121 TATTCCTTAGAAAGCAACAGCTAAAGAACTATTTCCTCAAGAACTGTGTGGCAAGTCAAGGGT 180  
 Db 12410 TATTCCTTAGAAAGCAACAGCTAAAGAACTATTTCCTCAAGAACTGTGTGGCAAGTCAAGGGT 12351

QY 181 AGTTAGTGATACACCGGTTATTCGGAGTCGGAGGAGAAAGTGTGTTCCGGCTGGACAAT 239  
 Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon

Db 12350 AGTTAGTGATACACCGGTTATATCGAGTGGCGAGAAAGTGTGGTTCGGCTGGACAAT 12292

RESULT 7  
AC095013/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 175648)

Holt, R.A., Evans, C.A., Gocayne, J.D., Adams, M.D., Krommiller, B., Tyler, D., Wan, K.H., Rogers, Y., An, H., Baldwin, D., Banzon, J., Beeson, K.Y., Busam, D.A., Carlson, J.W., Center, A., Champe, M., Davenport, L.B., Dietz, S.M., Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Dresnek, D., Farfan, D., Ferreria, S., Frisze, B., Galle, R.F., Garg, N.S., George, R.A., Gonzalez, M., Hock, J., Hoskins, R.A., Hostin, D., Howland, T.J., Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A., McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J., Pacleb, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B., Phouanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F., Stapleton, M., Strong, R., Svirskas, R., Tector, C., Williams, S.M., Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.

Sequencing of Drosophila chromosome 3R, region 83F-83P

Unpublished

2 (bases 1 to 175648)

Celniker, S.E., Adams, M.D., Krommiller, B., Tyler, D., Wan, K.H., Holt, R.A., Evans, C.A., Gocayne, J.D., Adams, M.D., Krommiller, B., Tyler, D., Wan, K.H., Rogers, Y., An, H., Baldwin, D., Banzon, J., Beeson, K.Y., Busam, D.A., Carlson, J.W., Center, A., Champe, M., Davenport, L.B., Dietz, S.M., Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Dresnek, D., Farfan, D., Ferreria, S., Frisze, B., Galle, R.F., Garg, N.S., George, R.A., Gonzalez, M., Hock, J., Hoskins, R.A., Hostin, D., Howland, T.J., Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A., McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J., Pacleb, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B., Phouanavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F., Stapleton, M., Strong, R., Svirskas, R., Tector, C., Williams, S.M., Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.

Direct Submission

Submitted (15-SEP-2001) Berkeley Drosophila Genome Project, MS 64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, US

Sequence submitted by:

Berkeley Drosophila Genome Project

Lawrence Berkeley National Laboratory, MS 64-121

Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://www.fruitfly.org/sequence/>) or send email to [bdgp@fruitfly.berkeley.edu](mailto:bdgp@fruitfly.berkeley.edu).

Location/Qualifiers

1. 175648

/organism="Drosophila melanogaster"

/mol\_type="genomic DNA"

/strain="y; cn bw sp"

/db\_xref="taxon:7227"

/chromosome="3R"

/map="83F-83P"

/clone="BACR19114 (D1348)"

/clone\_lib="RPCI-98 (Roswell Park Cancer Institute

Drosophila melanogaster BAC library, partial EcoRI in

pBACe3.6)"

ORIGIN

Query Match

Best Local Similarity

Matches 239; Conservative

0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCAGTCACGACTTGGTAAGCGCGAGCAGCAGCTCGTCTCGTCATCGCAACGGGAG 60

DB 144223 ATCAGTCACGACTTGGTAAGCGCGAGCAGCAGCTCGTCTCGTCATCGCAACGGGAG 144164

QY 61 TCGTGTGTTTTTCGGTTCGATACAGATAAAACCCAGCTCGATAGCCCTCGACCGTGCCTGATA 120

DB 144163 TCGTGTGTTTTTCGGTTCGATACAGATAAAACCCAGCTCGATAGCCCTCGACCGTGCCTGATA 144104

QY 121 TATCTTAGAAGCAACAGCTAAAGAACTATTTCAAGAACTGTGTGCAAGTGAAGGCT 180

DB 144103 TATCTTAGAAGCAACAGCTAAAGAACTATTTCAAGAACTGTGTGCAAGTGAAGGCT 144044

QY 181 AGTTAGTGATACACCGGTTATATCGAGTGGCGAGAAAGTGTGTTCGGCTGGACAAT 239

DB 144043 AGTTAGTGATACACCGGTTATATCGAGTGGCGAGAAAGTGTGTTCGGCTGGACAAT 143985

RESULT 8

AE003674/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Drosophila melanogaster (fruit fly)

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila

1 (bases 1 to 298020)

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D.,

Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F.,

George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N.,

Sutton, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X.,

Brandon, R.C., Rogers, Y.H., Blazer, R.G., Champe, M., Pfeiffer, B.D.,

Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor, G.L.,

Abriel, J.F., Agbayani, A., An, H.J., Andrews, Pfanckoch, C., Baldwin, D.,

Ballew, R.M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E.M.,

Beeson, K.Y., Benos, P.V., Berman, B.P., Bhandari, D., Bolshakov, S.,

Borkova, D., Botchan, M.R., Bouck, J., Brokstein, P., Brattier, P.,

Burtis, K.C., Busam, D.A., Butler, H., Cadieu, E., Center, A.,

Chandra, I., Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B.,

Dietz, S.M., Dodson, K., Doup, L.E., Downes, M., Mays, A.D., Dew, I.,

Dunkov, B.C., Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C.,

Ferriera, S., Fleischmann, W., Fosler, C., Gabrielson, A.E., Garg, N.S.,

QY 1 ATCAGTCACGACTTGGTAAGCGCGAGCAGCAGCTCGTCTCGTCATCGCAACGGGAG 60

DB 144223 ATCAGTCACGACTTGGTAAGCGCGAGCAGCAGCTCGTCTCGTCATCGCAACGGGAG 144164

QY 61 TCGTGTGTTTTTCGGTTCGATACAGATAAAACCCAGCTCGATAGCCCTCGACCGTGCCTGATA 120

DB 144163 TCGTGTGTTTTTCGGTTCGATACAGATAAAACCCAGCTCGATAGCCCTCGACCGTGCCTGATA 144104

QY 121 TATCTTAGAAGCAACAGCTAAAGAACTATTTCAAGAACTGTGTGCAAGTGAAGGCT 180

DB 144103 TATCTTAGAAGCAACAGCTAAAGAACTATTTCAAGAACTGTGTGCAAGTGAAGGCT 144044

QY 181 AGTTAGTGATACACCGGTTATATCGAGTGGCGAGAAAGTGTGTTCGGCTGGACAAT 239

DB 144043 AGTTAGTGATACACCGGTTATATCGAGTGGCGAGAAAGTGTGTTCGGCTGGACAAT 143985

RESULT 8

AE003674/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Drosophila melanogaster (fruit fly)

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila

1 (bases 1 to 298020)

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D.,

Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F.,

George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N.,

Sutton, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X.,

Brandon, R.C., Rogers, Y.H., Blazer, R.G., Champe, M., Pfeiffer, B.D.,

Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor, G.L.,

Abriel, J.F., Agbayani, A., An, H.J., Andrews, Pfanckoch, C., Baldwin, D.,

Ballew, R.M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E.M.,

Beeson, K.Y., Benos, P.V., Berman, B.P., Bhandari, D., Bolshakov, S.,

Borkova, D., Botchan, M.R., Bouck, J., Brokstein, P., Brattier, P.,

Burtis, K.C., Busam, D.A., Butler, H., Cadieu, E., Center, A.,

Chandra, I., Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B.,

Dietz, S.M., Dodson, K., Doup, L.E., Downes, M., Mays, A.D., Dew, I.,

Dunkov, B.C., Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C.,

Ferriera, S., Fleischmann, W., Fosler, C., Gabrielson, A.E., Garg, N.S.,

Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z.,

Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J.,

Hernandez, J.R., Hock, J., Hostin, D., Houston, K.A., Howland, T.J.,

Wei, M.H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z.,

QY 1 ATCAGTCACGACTTGGTAAGCGCGAGCAGCAGCTCGTCTCGTCATCGCAACGGGAG 60

DB 144223 ATCAGTCACGACTTGGTAAGCGCGAGCAGCAGCTCGTCTCGTCATCGCAACGGGAG 144164

QY 61 TCGTGTGTTTTTCGGTTCGATACAGATAAAACCCAGCTCGATAGCCCTCGACCGTGCCTGATA 120

DB 144163 TCGTGTGTTTTTCGGTTCGATACAGATAAAACCCAGCTCGATAGCCCTCGACCGTGCCTGATA 144104

QY 121 TATCTTAGAAGCAACAGCTAAAGAACTATTTCAAGAACTGTGTGCAAGTGAAGGCT 180

DB 144103 TATCTTAGAAGCAACAGCTAAAGAACTATTTCAAGAACTGTGTGCAAGTGAAGGCT 144044

QY 181 AGTTAGTGATACACCGGTTATATCGAGTGGCGAGAAAGTGTGTTCGGCTGGACAAT 239

DB 144043 AGTTAGTGATACACCGGTTATATCGAGTGGCGAGAAAGTGTGTTCGGCTGGACAAT 143985

RESULT 8

AE003674/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Drosophila melanogaster (fruit fly)

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila

1 (bases 1 to 298020)

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D.,

Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F.,

George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N.,

Sutton, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X.,

Brandon, R.C., Rogers, Y.H., Blazer, R.G., Champe, M., Pfeiffer, B.D.,

Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor, G.L.,

Abriel, J.F., Agbayani, A., An, H.J., Andrews, Pfanckoch, C., Baldwin, D.,

Ballew, R.M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E.M.,

Beeson, K.Y., Benos, P.V., Berman, B.P., Bhandari, D., Bolshakov, S.,

Borkova, D., Botchan, M.R., Bouck, J., Brokstein, P., Brattier, P.,

Burtis, K.C., Busam, D.A., Butler, H., Cadieu, E., Center, A.,

Chandra, I., Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B.,

Dietz, S.M., Dodson, K., Doup, L.E., Downes, M., Mays, A.D., Dew, I.,

Dunkov, B.C., Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C.,

Ferriera, S., Fleischmann, W., Fosler, C., Gabrielson, A.E., Garg, N.S.,

Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z.,

Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J.,

Hernandez, J.R., Hock, J., Hostin, D., Houston, K.A., Howland, T.J.,

Wei, M.H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z.,



```

/note="last curated on Thu Jan 03 10:58:20 PST 2002"
/map="83F1-83F1"
/db_xref="FLYBASE:FBgn0037449"
47311..48339
/locus_tag="CG15185"

Query Match      100.0%; Score 239; DB 3; Length 298020;
Best Local Similarity 100.0%; Pred. No. 61e-66;
Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATCAGTCACGACTGGTAAAGCGCGAGCAGCAGCAGTCTGTCGTCATCGCCAAAGCGGAG 60
Db 198367 ATCAGTCACGACTGGTAAAGCGCGAGCAGCAGCAGTCTGTCGTCATCGCCAAAGCGGAG 198308

Qy 61 TCGTGTCTTTTCGGTTCGATACAGATAAACCACGTCATAGCCCTCAGCGTCGCGTAA 120
Db 198307 TCGTGTCTTTTCGGTTCGATACAGATAAACCACGTCATAGCCCTCAGCGTCGCGTAA 198248

Qy 121 TATCTTTAGAAAGCAACAGCTAAAGAACTATTTCAAGAACTGTGTGCAAGTGAAGGCT 180
Db 198247 TATCTTTAGAAAGCAACAGCTAAAGAACTATTTCAAGAACTGTGTGCAAGTGAAGGCT 198188

Qy 181 AGTAGTCATACACCGTTATATCGAGTGGCGAGAAAGTGTGTCGCGTGCACAAAT 239
Db 198187 AGTAGTCATACACCGTTATATCGAGTGGCGAGAAAGTGTGTCGCGTGCACAAAT 198129

RESULT 9
AC106665 237630 bp DNA linear HTG 13-MAY-2003
LOCUS Rattus norvegicus clone CH230-102113, WORKING DRAFT SEQUENCE, 3
DEFINITION Rattus norvegicus (Norway rat)
ACCESSION AC106665
VERSION 4 GI:30578557
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

1 (bases 1 to 237630)
Muzny,D,Marie., Metzker,M,Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Drapet,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,M.,
Gunaratne,P., Haaland,W., Hamill,C., Hamilton,N., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hollnandez,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovat,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokelimeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,

```

```

Pasternak,S., Paul,H., Perez,A., Perez,L., Pfankoch,C.,
Plopper,F., Poindeexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajic,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Vallas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,O., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wlezyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,X., Zhou,S., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.

Direct Submission
Unpublished
2 (bases 1 to 237630)
Worley,K.C.
Direct Submission
Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 237630)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 13, 2003 this sequence version replaced gi:22857405.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GLT2
Center clone name: CH230-102113
----- Summary Statistics
Assembly program: Atlas 3.0:
Consensus quality: 229128 bases at least Q40
Consensus quality: 231415 bases at least Q30
Consensus quality: 233051 bases at least Q20
Estimated insert size: 240583; sum-of-contigs estimation
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1
* 234955 contig of 234954 bp in length
* 235054 gap of unknown length
* 235055 contig of 1387 bp in length

```

```
* 236442 236541: gap of unknown length
* 236542 237630: contig of 1089 bp in length.
FEATURES
    source
        1..237630
            /organism="Rattus norvegicus"
            /mol_type="genomic DNA"
            /db_xref="taxon:10116"
            /clone="CH230-102113"
            1..1505
            /note="wgs_contig"
            2205..4409
            /note="wgs_contig"
ORIGIN
Query Match      15.3%; Score 36.6; DB 2; Length 237630;
Best Local Similarity 58.9%; Pred. No. 4.1;
Matches 63; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
QY 120 ATATTCTTGAAGCAACAGCTAAAGAACTATTTCACAGACTGTGTGCAAGTGAAGG 179
Db 215052 ATAGTTTGGAAACCAAGCACTAGATAAATATTTATCAACAGAGTGAATGCTAGAA 215111
QY 180 TAGTTAGTGATACACCGGTTATATCGAGTGGCGAGAAAGTGTGTT 226
Db 215112 TAGAGATTAGATGGTGTGTTTCTGATGACAGGAGTTCAGCTT 215158
RESULT 10
AC131194
LOCUS
DEFINITION
AC131194 Rattus norvegicus clone CH230-85M5, WORKING DRAFT SEQUENCE.
AC131194
VERSION
AC131194.3 GT:30522034
KEYWORDS
HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE
Rattus norvegicus
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 239015)
Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Caesar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Kocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregiorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorenshewa,L., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K.,
Nwakoelameh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
```

```
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Relly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.D.,
Sanders,J., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 239015)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (18-AUG-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 239015)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GGBH
Center clone name: CH230-85M5
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 224202 bases at least Q40
Consensus quality: 226474 bases at least Q30
Consensus quality: 228026 bases at least Q20
Estimated insert size: 235880; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases, sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 239015: contig of 239015 bp in length.
FEATURES
    Location/Qualifiers
    1..239015
```



```

/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-85MS"
1. .1575
/Note="wgs end extension
clone end:Sp6"
10013. .10837
/Note="clone boundary
clone end:Sp6
site:EcORI
end_sequence:BH305681"
complement(235727..236614)
/Note="clone_boundary
clone end:T7
site:EcORI
end_sequence:BH305680"

ORIGIN
Query Match 15.3%; Score 36.6; DB 2; Length 239015;
Best Local Similarity 58.9%; Pred No. 4.1;
Matches 63; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

Qy 120 ATATTCTTAGAAGCAACACGACTAAGAACTATTTCACAGAACTGTGTGCGCAAGTCAAGGG 179
Dy 14324 ATAGTTTGGAAACCAACGCAACTAGATAATATTTTATCAACAGAGTGGAATGCTAGAA 14383

Qy 180 TAGTGTAGTACACCGGTTATTCGAGTGGCGAGAAAGTGTGTT 226
Dy 14384 TAGAGCATTAGATGCTGTGTTTCTGAATGACAAAGGAGTTCAGCTT 14430

RESULT 11
AC096362/c
LOCUS
DEFINITION Rattus norvegicus clone CH230-53N24, WORKING DRAFT SEQUENCE.
ACCESSION AC096362
VERSION AC096362.6 GI:30521829
KEYWORDS HTGS PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 282019)
Muzny,D,Maris,, Mezker,M, Lee,, Abramson,S,, Adams,C,, Alder,J,,
Allen,C,, Allen,H,, Alsbrooks,S,, Amin,A,, Anguiano,D,,
Anyalebechi,V,, Aoyagi,A,, Ayodeji,M,, Baca,E,, Baden,H,,
Baldwin,D,, Bandaranaike,D,, Barber,M,, Barnstead,M,, Benahmed,F,,
Biswal,K,, Blair,J,, Blankenburg,K,, Blyth,P,, Brown,M,,
Bryant,N,, Buhay,C,, Burch,P,, Burrell,K,, Calderon,E,,
Cardenas,V,, Carter,K,, Cavazos,I,, Ceasar,H,, Center,A,,
Chacko,J,, Chavez,D,, Chen,G,, Chen,R,, Chen,Y,, Chen,Z,, Chu,J,,
Cleveland,C,, Cockrell,R,, Cox,C,, Coyle,M,, Cree,A,, D'Souza,L,,
Davila,M,L,, Davis,C,, Davy-Carroll,L,, De Anda,C,, Dederich,D,,
Delgado,O,, Denison,S,, Deramo,C,, Ding,Y,, Dinh,H,, Divya,K,,
Draper,H,, Dugan-Rocha,S,, Dunn,A,, Durbin,K,, Duval,B,, Eaves,K,,
Egan,A,, Escotto,M,, Eugene,C,, Evans,C,A,, Falls,T,, Fan,G,,
Fernandez,S,, Finley,M,, Flagg,N,, Forbes,I,, Foster,M,, Foster,P,,
Fraser,C,M,, Gabisi,A,, Ganta,R,, Garcia,A,, Garner,T,, Garza,M,,
Gebregeorgis,E,, Geer,K,, Gill,R,, Grady,M,, Guerra,W,, Guevara,W,,
Gunaratne,P,, Haaland,W,, Hamil,C,, Hamilton,C,, Hamilton,K,,
Harvey,Y,, Havlak,P,, Hawes,A,, Henderson,N,, Hernandez,J,,
Hernandez,R,, Hines,S,, Hladun,S,L,, Hodgson,A,, Hogues,M,,
Hollins,B,, Howells,S,, Hulyk,S,, Hume,J,, Idlebird,D,, Jackson,A,,
Jackson,L,, Jacob,H,, Jiang,H,, Johnson,B,, Johnson,R,, Jolivet,A,,
Karpachy,S,, Kelly,S,, Kelly,S,, Khan,Z,, King,L,, Kovar,C,,
Kowis,C,, Kraft,C,L,, Lebow,H,, Levan,J,, Lewis,L,, Li,Z,, Liu,J,,
Liu,J,, Liu,W,, Liu,Y,, London,P,, Longacre,S,, Lopez,J,,
Lorensuhewa,L,, Loulsegad,H,, Lozad,R,J,, Lu,X,, Ma,J,,
Maheshwari,M,, Mahindartine,M,, Mahmoud,M,, Malloy,K,, Mangum,A,,
Mangum,B,, Mapua,P,, Martin,K,, Martin,R,, Martinez,E,,
Mawhiney,S,, McLeod,M,P,, McNeill,T,Z,, Meenen,E,,

```

Milosavljevic,A., Miner,G., Minja,E., Montenayor,J., Moore,S.,  
 Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,  
 Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,  
 Nwackemehe,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,  
 Pasternak,S., Paul,H., Perez,A., Perez,L., Pfankoch,C.,  
 Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.,  
 Puazo,M., Quiroz,J., Rachin,E., Reeves,K., Regier,M.A., Reigh,R.,  
 Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,  
 Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,  
 Sanders,M., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,  
 Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajls,D.,  
 Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,  
 Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,  
 Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,  
 Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,  
 Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,  
 Williams,G., Willson,R., Wlezyk,R., Wooden,H., Worley,K.,  
 Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,  
 Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von  
 Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,  
 Weinstock,G. and Gibbs,R.A.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 282019)  
 Worley,K.C.  
 Direct Submission  
 Submitted (17-SEP-2001) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 282019)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (10-MAY-2003) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On May 10, 2003 this sequence version replaced gi:2271585.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold,  
 individual sequence contigs are ordered and oriented, and separated  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature  
 table.  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: http://www.hgsc.bcm.tmc.edu/  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GEZO  
 Center clone name: CH230-53N24  
 ----- Summary Statistics  
 Assembly program: Atlas 3.0;  
 Consensus quality: 236414 bases at least Q40  
 Consensus quality: 239934 bases at least Q30  
 Consensus quality: 242193 bases at least Q20  
 Estimated insert size: 249348; sum-of-contigs estimation  
 Quality coverage: 7x in Q20 bases; sum-of-contigs estimation  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 1 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* provided by the submitter.  
 \* This sequence will be replaced

\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 282019: contig of 282019 bp in length.

FEATURES  
source  
1..282019  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-53N24"  
4444..5560  
/note="wgs\_contig"  
misc\_feature  
279586..280897  
/note="wgs\_contig"  
misc\_feature  
280948..282019  
/note="wgs\_contig"

ORIGIN

Query Match 15.3%; Score 36.6; DB 2; Length 282019;  
Best Local Similarity 58.9%; Pred. No. 4.2;  
Matches 63; Conservative 0; Mismatches 44; Indels 0; Gaps 0;  
QY 120 ATATTCTTAGAAGCAACAGCTAAAGAACTATTTCAGAACTGTGTGCAAGTGAAGG 179  
Db 258385 ATAGTTTGGAAACCAAGCAACTAGATAAATTTTATCAACAGAGTGAATGCTAGAA 258326  
QY 180 TAGTTAGTACATACACCGTTATATCGGAGTGCAGAGAAAGTGTGTT 226  
Db 258325 TAGAGCAATTAGATGCTGTGTGTTTCTGAATGCAAGGAGTTCAGCTT 258279

RESULT 12

LOCUS  
AE008718 20938 bp DNA linear BCT 23-APR-2003  
DEFINITION  
Salmonella typhimurium LT2, section 26 of 220 of the complete genome.  
ACCESSION  
AE008718 AB006468  
VERSION  
AE008718.1 GI:16418985

KEYWORDS

Salmonella typhimurium LT2

SOURCE

Salmonella typhimurium LT2

Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

Enterobacteriaceae; Salmonella.

1 (bases 1 to 20938)

McClelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W., Du,F.,

Latreille,P., Courtney,L., Porwollik,S., Ali,J., Dante,M., Du,F.,

Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A.,

Grewal,N., Mulvaney,E., Ryan,E., Sun,H., Florea,L., Miller,W.,

Stonking,T., Nhan,M., Waterston,R. and Wilson,R.K.

Complete genome sequence of Salmonella enterica serovar Typhimurium

LT2

Nature 413 (6858), 852-856 (2001)

21534948

11677609

2 (bases 1 to 20938)

The Salmonella typhimurium Genome Sequencing Project

Direct Submission

Submitted (29-MAR-2001) Genome Sequencing Center, Department of

Genetics, Washington University School of Medicine, 4444 Forest

Park Boulevard, St. Louis, MO 63108, USA

COMMENT Supported by NIH grant 5U 01 AI43283

Coding sequences below are predicted from manually evaluated

computer analysis, using similarity information and the programs;

GLIMMER; <http://www.tigr.org/softlab/glimmer/glimmer.html> and

GeneMark; <http://opal.biology.gatech.edu/GeneMark/>

EC numbers were kindly provided by Junko Yabuzaki and the Kyoto

Encyclopedia of Genes and Genomes; <http://www.genome.ad.jp/kegg/>,

and Pedro Romero and Peter Karp at Ecocyc;

<http://ecocyc.PangeaSystems.com/ecocyc/>

The analyses of ribosome binding sites and promoter binding sites

were kindly provided by Heladia Salgado, Julio Collado-Vides and  
RegulonDB;  
[http://kinich.cifn.unam.mx:8850/db/regulondb\\_intro.frameset](http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset)

This sequence was finished as follows unless otherwise noted: all  
regions were double stranded, sequenced with an alternate  
chemistries or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by sequence  
from more than one ml3 subclone.

FEATURES

source  
1..20938  
/organism="Salmonella typhimurium LT2"  
/mol\_type="genomic DNA"  
/strain="LT2; SGSC 1412; ATCC 700720"  
/db\_xref="ATCC:700720"  
/db\_xref="taxon:99287"  
/note="LT2"  
90..3452  
/gene="aefa"  
/note="synonym: STM0478"  
90..3452  
/gene="aefa"  
/note="similar to E. coli putative alpha helix protein  
(AAC73567.1); Blastp hit to AAC73567.1 (1120 aa), 89%  
identity in aa 1 - 1117"  
/codon\_start=1  
/transl\_table=11  
/product="putative small-conductance mechanosensitive  
channel"  
/protein\_id="AAL19432.1"  
/db\_xref="GI:16418986"  
/translation="MTMLQYKRSQHFFVITISVLIIILSCSLAFARGQTNGDLEPK  
ADVQQLDNLNRKQDLSAQKLVQDLDLDTLATEKIERVEETVQLKQKVAQAEKX  
RQATAALNALSDVNDDEMRKLSALSRLQLELRVAQVDDDLQNSQDLAAYNSQVLS  
LQTPERVQNAVMTASQIQIIRNRDLGNVGEALRPSQVLLQQAQLNALNAIDQQ  
RKSLEGTVLQDTQKQDYVTANSRLHQILQLEAVNSKRLTLTEKTAQEAISPD  
ETARTQANPLVKQELDIINHLSQRLIVATENGMLMCONIKVKNWLDRAQLQSERNIKE  
QIAYLGSLLSRILYQQQQLPSADELDMTNRIADLRLEQFELNQORDALFQSDAF  
VYKLEEGHSTVNDVHDALLOVEMRRELLDQLNQLGNQLMAINLQVNNQQLMSV  
SKNLKALITQIFWYNSRPMWDMDKAPFQTLKEQPSAMKITVNWQKAPVFIAPL  
SKGLLIAGLIRWKLKQYQKLAAGVSLRNSQNTPKALIDILRALPVLCL  
ILALGLIILLTQMLNISDLIWFASKLMLVFLGKWLKEGVVIRHFGMPAQLTS  
HWRQIVRISIALPLSHFWSVVAELSPNLMDVLGQAVIFLNLIETVLIVLWLLYQ  
WDEKSHGIRLVYTLISIPVALMVLATGYFYTLRAGIEWITVILVILWLLYQ  
TVLRGSLVAARRIARNRALRQNLVKGAGARQEQEPTIALEQINQOQLTILMLYQ  
LALFQVFWFAWSDLITVFSYDLSITLWHYSGEAGAAVVKSVTSGSLGVSQDLQW  
LAAALSVGLGFLQEIFGNFVSGLIILFERPVIRIGTIVTIGTYSKIRIRATIT  
DPRKEVIIPKAFVYERLINWSDTTRILVIRLGVAYGSDLEKVKRVLLOAAHEHP  
KVMHDPKPAVPTTFTGASTLDHELRLVYRELDRSHVTDELNRALDRLCRENDINIAF  
NQLFVHLNNAKGDEVTEVKRLDNGGDLAPTAS"  
complement(3494..4435)  
/gene="STM0479"  
complement(3494..4429)  
/gene="STM0479"  
/note="similar to E. coli orf, hypothetical protein  
(AAC75365.1); Blastp hit to AAC75365.1 (296 aa), 49%  
identity in aa 7 - 296"  
/codon\_start=1  
/transl\_table=11  
/product="putative transposase"  
/protein\_id="AAL19433.1"  
/db\_xref="GI:16418987"  
/translation="MERIPTPHDAVFRQMLMKQKVARDFLAHMPEDFLAICDLDSL  
KLESGFVEDNLRYSRITLLSYLTHQHGPGYVIALIEHQSKSDRLMAFKRMVIAIAM  
QRHLDAHDHTELVYFVILFHGFPSPFVSLNWNHNFVKNWMAKLSVDFALVYLKA  
MDPNQLQHRRIAMLELLQKHIRQDLSELLDPLITLTDHLTDQSLSDVLYNMLKA  
GNAAPGALIROLAQCAPQYKQELMTIAEWLEKGRTEGLQGLQKLGLEQQAQREGA  
EAAATARKMLANGPGLASVYTGITPELSTLSH"  
complement(4430..4435)  
/gene="STM0479"  
/note="putative RBS for STM0479; RegulonDB: STM5IH000821"

gene

CDS

RBS

```

gene
complement(4499..4677)
/ gene="ybam"
/ note="synonym: STM0480"
CDS
complement(4499..4666)
/ gene="ybam"
/ note="similar to E. coli orf, hypothetical protein
(AAC73568.1); Blastp hit to AAC73568.1 (53 aa), 80%
identity in aa 1 - 49"
/ codon_start=1
/ transl_table=11
/ product="putative inner membrane protein"
/ protein_id="AAL19434.1"
/ db_xref="GI:16418988"
/ translation="MSLENAPDEVKLVLDVLLLENRLPARTVLRALVIVRDYENK
LKSTEDDSQTE"
RBS
complement(4672..4677)
/ gene="ybam"
/ note="putative RBS for ybam; RegulonDB: STMSIH000822"
gene
complement(4680..5195)
/ gene="pric"
/ note="synonym: STM0481"
CDS
complement(4680..5195)
/ gene="pric"
/ note="similar to E. coli primosomal replication protein
N', (AAC73569.1); Blastp hit to AAC73569.1 (175 aa), 73%
identity in aa 5 - 175"
/ codon_start=1
/ transl_table=11
/ product="primosomal replication protein N"
/ protein_id="AAL19435.1"
/ db_xref="GI:16418989"
/ translation="MLLQTLERLTLRQRCAPLAQHATLSARFDRHLFRFTSLTGLQ
YLERAGNVALROAVKEQLPOVAMLAHLASQLEALSRETAWSLRQWDAAPGLG
PWQRRTQHQFERLLANTQERKIRLAQTGLVEQQLQKEVIEYERLARCRHALE
KIENVLARLT"
5276..5653
/ gene="yban"
/ note="synonym: STM0482"
5276..5653
/ gene="yban"
/ note="similar to E. coli putative gene 58 (AAC73570.1);
Blastp hit to AAC73570.1 (125 aa), 84% identity in aa 1 -
124"
/ codon_start=1
/ transl_table=11
/ product="putative phage gene 58"
/ protein_id="AAL19436.1"
/ db_xref="GI:16418990"
/ translation="MQRTILIIIGWLAIVLGTGLWVLPPLPTTPFILLAAWCFARSSP
RFHAWLLYRSWFGYLRHQRYRAMPPGAKPRATALLITFGISLWLVNMMVVRVLL
VILACLLIFMWRIPVDEKQQR"
5793..6357
/ gene="apt"
/ note="synonym: STM0483"
5793..5798
/ gene="apt"
/ note="putative RBS for apt; RegulonDB: STMSIH000823"
5806..6357
/ gene="apt"
/ EC number="2.4.2.7"
/ note="similar to E. coli adenine
phosphoribosyltransferase (AAC73571.1); Blastp hit to
AAC73571.1 (183 aa), 95% identity in aa 1 - 183"
/ codon_start=1
/ transl_table=11
/ product="adenine phosphoribosyltransferase"
/ protein_id="AAL19437.1"
/ db_xref="GI:16418991"
/ translation="MTATACQLQFLKNSIKSIQDPKPGILFRDVTSLLEDPKAYALS
IELLYRYKAGITKVGTGARGFLGAPVALGLGVFPVRPKRLPRETIAETVEL
EYGTQQLIEHVDIAKPGDNLVVDLATGTGTATVKLIRRLGGKVTDAAFINLFD
LGGEQLKQGITCYSIVPPFGH"
6471..8399

```

```

/ gene="dnaX"
/ note="synonym: STM0484"
6471..8399
/ gene="dnaX"
/ EC number="2.7.7.7"
/ note="DNA elongation factor III; DNA polymerase III
subunit TAU (contains: DNA polymerase III subunit gamma).
(SW:DP3X_SALTY)"
/ codon_start=1
/ transl_table=11
/ product="DNA polymerase III, tau and gamma subunits"
/ protein_id="AAL19438.1"
/ db_xref="GI:16418992"
Query Match 14.6%; Score 34.8; DB 1; Length 20938;
Best Local Similarity 52.0%; Pred. No. 10;
Matches 78; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
QY 23 GCAGGAGCAGCGTCGTCGTCATCGCCACAGGGAGTCGTGTTTTTCGGTTCGATACA 82
Db 18666 GCAGCGAGCAATACGCGCCAAAACCTTCGCGCATTCGCAGCAGTACCGCCGCTCAG 18725
QY 83 GATAAAACCCAGCTCGATAGCCCTCGACCGTCGGTATATCTTAGAAACACACAGCT 142
Db 18726 GGTAAACACATGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 142
QY 143 AAAGAACTATTTCAGAACTGTGTGGCAAG 172
Db 18786 GACGACCGGTTCGACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 18815
RESULT 13
AC099723/c 178993 bp DNA linear HTG 19-OCT-2002
LOCUS Mus musculus clone RP23-369P21, WORKING DRAFT SEQUENCE, 8 unordered
DEFINITION pieces.
AC099723
VERSION AC099723.2 GI:24137577
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1. (bases 1 to 178993)
Mus musculus, clone RP23-369P21
Unpublished
2. (bases 1 to 178993)
Birren, B., Linton, B., Nussbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Choepel, Y., Collangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,
Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,
Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J.,
Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (18-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

```

## REFERENCE

3 (bases 1 to 178993)  
 Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,  
 Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,  
 Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,  
 Cooke,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
 Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,  
 Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,  
 Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
 Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,  
 Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,  
 Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,  
 Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,  
 Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,  
 Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,  
 Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,  
 Seaman,S., Sever,P., Smith,C., Spencer,B., Stange-Thomann,N.,  
 Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,  
 Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,  
 Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (19-OCT-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Oct 19, 2002 this sequence version replaced gi:16974222.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RX/RepeatMasker.html

## TITLE

## JOURNAL

## COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L18019

Center clone name: 369 p 21

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 176482 bases at least Q40

Consensus quality: 177487 bases at least Q30

Consensus quality: 177924 bases at least Q20

Insert size: 183000; agarose-ff

Insert size: 178293; sum-of-contigs

Quality coverage: 10.8 in Q20 bases; agarose-ff

Quality coverage: 11.1 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 8 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 8673: contig of 8673 bp in length  
 \* 8674 8773: gap of 100 bp  
 \* 8774 9858: contig of 1085 bp in length  
 \* 9859 9958: gap of 100 bp  
 \* 9959 19220: contig of 9262 bp in length  
 \* 19221 19320: gap of 100 bp  
 \* 19321 29525: contig of 10205 bp in length  
 \* 29526 29625: gap of 100 bp  
 \* 29626 95507: contig of 6582 bp in length  
 \* 95508 95607: gap of 100 bp  
 \* 95608 128597: contig of 32990 bp in length  
 \* 128598 128697: gap of 100 bp  
 \* 128698 172325: contig of 43628 bp in length  
 \* 172326 172426: gap of 100 bp  
 \* 172426 178993: contig of 6568 bp in length.

## FEATURES

## source

1. .178993  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"

/db\_xref="taxon:10090"  
 /clone="RP23-369P21"  
 /clone\_lib="RPC1-23 Female Mouse BAC"  
 1. .8673  
 /note="assembly\_fragment  
 clone\_end:SP6  
 vector\_side:left"  
 8774. .9858  
 /note="assembly\_fragment"  
 9959. .19220  
 /note="assembly\_fragment"  
 19321. .29525  
 /note="assembly\_fragment"  
 29626. .95507  
 /note="assembly\_fragment"  
 95608. .128597  
 /note="assembly\_fragment"  
 128698. .172325  
 /note="assembly\_fragment"  
 172426. .178993  
 /note="assembly\_fragment  
 clone\_end:17  
 vector\_side:right"

## ORIGIN

Query Match 14.2%; Score 34; DB 2; Length 178993;  
 Best Local Similarity 51.3%; Pred No. 27;  
 Matches 79; Conservative 0; Mismatches 75; Indels 0; Gaps 0;  
 QY 83 GATAAAACCCAGCTCGATAGCCCTCGAGCTCGCTTAATATTTCTTTAGAAAGCAACAGCT 142  
 Db 115682 GCTTAACACAGAGCCATCTCCCGAGGCTCTGGAAGTAAACTTTTAAACCACTTTT 115623  
 QY 143 AAAGAACTATTTCAGAACTGTGTGGCAAGTGAAGGTAGTTAGTGATACACCGGTATA 202  
 Db 115622 ATAGTTCTTTTAAATAATATGTTATTTCAATTATGTTGGGAGGAGCTCAATGCTTA 11563  
 QY 203 TCGGAGTGGCGAGAAAGTGTGTTCCGGCTGCAC 236  
 Db 115562 GAAGCTAGAGAGGATGTGACCCCGGGGCTGTC 115529

## RESULT 14

## AC113984

## LOCUS

## DEFINITION

## AC113984

## AC113984

## AC113984

## AC113984

## AC113984

## AC113984

## AC113984

## AC113984

## AC113984

## AC113984

## AC113984

## AC113984

## AC113984

## AC113984

## AC113984

## AC113984

## AC113984

## AC113984

## AC113984

## AC113984

## AC113984

## AC113984

## AC113984

## AC113984

## AC113984

## AC113984

## AC113984

## AC113984

## AC113984

## AC113984

Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,  
 Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,  
 Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S.,  
 Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
 Straus, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
 Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,  
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,  
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

# TITLE

## JOURNAL

### REFERENCE

### AUTHORS

Submitted (06-MAR-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 193636)

Birken, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,  
 Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhalter, B.,  
 Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,  
 Cook, A., Cooke, P., DeRellano, K., Dewar, K., Diaz, J. S., Dodge, S.,  
 Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,  
 Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hafez, N.,  
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
 Kanat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,  
 Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J.,  
 Matthews, C., McCarthy, M., Meldrum, J., Meneus, L., Mihova, T.,  
 Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,  
 Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,  
 Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R.,  
 Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupback, R.,  
 Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,  
 Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K.,  
 Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X.,  
 Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

# TITLE

## JOURNAL

### COMMENT

Submitted (21-OCT-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Oct 21, 2002 this sequence version replaced g1:21327615.  
 All repeats were identified using RepeatMasker:  
 Smit, A. F. A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WBFR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence.submissions@genome.wi.mit.edu](mailto:sequence.submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L24343

Center clone name: 97\_D\_3

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 189906 bases at least Q40

Consensus quality: 191355 bases at least Q30

Consensus quality: 191926 bases at least Q20

Insert size: 168000; agarose-fp

Quality coverage: 9.5 in Q20 bases; sum-of-contigs

Quality coverage: 8.3 in Q20 bases; sum-of-contigs

-----

NOTE: This is a 'working draft' sequence. It currently  
 consists of 13 contigs. Gaps between the contigs  
 are represented as runs of N. The order of the pieces  
 is believed to be correct as given, however the sizes  
 of the gaps between them are based on estimates that have  
 been provided by the submitter.

This sequence will be replaced

by the finished sequence as soon as it is available and

the accession number will be preserved.

1 1188: contig of 1188 bp in length

1189 1288: gap of 100 bp

1289 3601: contig of 2313 bp in length

3602 3701: gap of 100 bp

3702 6708: contig of 3007 bp in length

6709 6808: gap of 100 bp

6809 8666: contig of 1858 bp in length

\* 8667 8766: gap of 100 bp  
 \* 8767 13690: contig of 4924 bp in length  
 \* 13691 13790: gap of 100 bp  
 \* 13791 62565: contig of 48775 bp in length  
 \* 62566 62665: gap of 100 bp  
 \* 62666 65008: contig of 3843 bp in length  
 \* 65009 66008: gap of 100 bp  
 \* 66009 74500: contig of 7892 bp in length  
 \* 74501 74600: gap of 100 bp  
 \* 74601 97683: contig of 23083 bp in length  
 \* 97684 97783: gap of 100 bp  
 \* 97784 134290: contig of 36507 bp in length  
 \* 134291 134390: gap of 100 bp  
 \* 134391 158667: contig of 24277 bp in length  
 \* 158668 158767: gap of 100 bp  
 \* 158768 189834: contig of 31067 bp in length  
 \* 189835 189934: gap of 100 bp  
 \* 189935 193636: contig of 3702 bp in length.

### FEATURES

#### source

1..193636  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /clone="RP24-97D3"  
 /clone\_lib="RPCI-24 Male Mouse BAC"

#### misc\_feature

1..1188  
 /note="assembly\_fragment"

#### misc\_feature

1289..3601  
 /note="assembly\_fragment"

#### misc\_feature

3702..6708  
 /note="assembly\_fragment"

#### misc\_feature

6809..8666  
 /note="assembly\_fragment"

#### misc\_feature

8767..13690  
 /note="assembly\_fragment"

#### misc\_feature

13791..62565  
 /note="assembly\_fragment"

#### misc\_feature

62666..66508  
 /note="assembly\_fragment"

#### misc\_feature

66609..74500  
 /note="assembly\_fragment"

#### misc\_feature

74601..97683  
 /note="assembly\_fragment"

#### misc\_feature

97784..134290  
 /note="assembly\_fragment"

#### misc\_feature

134391..158667  
 /note="assembly\_fragment"

#### misc\_feature

158768..189834  
 /note="assembly\_fragment"

#### misc\_feature

189935..193636  
 /note="assembly\_fragment"

clone\_end:T7  
 vector\_side:right

### ORIGIN

Query Match 14.0%; Score 33.4; DB 2; Length 193636;  
 Best Local Similarity 54.5%; Pred. No. 43;  
 Matches 67; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 50 GCCAACGGAGTCGTGTTTTTCGTTACAGATATAACCCACGTCGATAGCCCTCGA 109

Db 3256 GGCACGGGGCGCTGTGATTTCTTTAGTTTCAGAGCCTCAGACGTCCAAATCCCTGTGA 3315

QY 110 CCCTCGCGTAATATTCCTGAAAGCAACAGCTTAAAGAACTATTTCAAGAACTGTGTGGC 169

Db 3316 CTCTGAGTTGTTATTTCTGGGCCCAAAAAAAGAGGCTGAGTGGC 3375

QY 170 AAG 172

Db 3376 CAG 3378

RESULT 15

Mon Sep 20 12:59:40 2004

## AC119849

LOCUS AC119849 229102 bp DNA linear HTG 22-FEB-2003  
 DEFINITION Mus musculus clone RP23-7404, WORKING DRAFT SEQUENCE, 12 ordered pieces.

## AC119849

VERSION AC119849.2 GI:28467231  
 KEYWORDS HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.  
 SOURCE Mus musculus (house mouse)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE

1 (bases 1 to 229102)

## AUTHORS

Birren,B., Nusbaum,C. and Lander,B.

## TITLE

Mus musculus, clone RP23-7404

## JOURNAL

Unpublished

## REFERENCE

2 (bases 1 to 229102)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano A., Chang,J., Chararo,B., Choepel,X., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gird,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,K., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

## Direct Submission

## TITLE

Submitted (02-MAY-2002) Whitehead Institute/MIT Center for Genome

## JOURNAL

Research, 320 Charles Street, Cambridge, MA 02141, USA

## REFERENCE

3 (bases 1 to 229102)

## AUTHORS

Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,X., Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-pierre,N., Hafez,N., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., Maclean,C., Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

## Direct Submission

## TITLE

Submitted (22-FEB-2003) Whitehead Institute/MIT Center for Genome

## JOURNAL

Research, 320 Charles Street, Cambridge, MA 02141, USA

## COMMENT

On Feb 22, 2003 this sequence version replaced gi:20389488.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Contact: sequence submissions@genome.wi.mit.edu

----- Project Information

Center project name: 122358

Center clone name: 74\_O\_4

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 228642 bases at least Q40

Consensus quality: 227526 bases at least Q30

Consensus quality: 227772 bases at least Q20

Insert size: 212000; agarose-fp

Insert size: 228002; sum-of-contigs

Quality coverage: 11.5 in Q20 bases; agarose-fp

Quality coverage: 10.7 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently consists of 12 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

\* This sequence will be replaced

\* by the finished sequence as soon as it is available and \* the accession number will be preserved.

\* 1 12780: contig of 12780 bp in length

\* 12781 12880: gap of 100 bp

\* 12881 13370: contig of 490 bp in length

\* 13371 13470: gap of 100 bp

\* 13471 14758: contig of 1288 bp in length

\* 14759 14858: gap of 100 bp

\* 14859 17963: contig of 3105 bp in length

\* 17964 18063: gap of 100 bp

\* 18064 20876: contig of 2813 bp in length

\* 20877 27426: gap of 100 bp

\* 27427 27526: gap of 100 bp

\* 27527 41619: contig of 14093 bp in length

\* 41620 41720: contig of 56421 bp in length

\* 41720 98240: gap of 100 bp

\* 98241 120308: contig of 22068 bp in length

\* 120309 120408: gap of 100 bp

\* 120409 160843: contig of 40435 bp in length

\* 160844 160943: gap of 100 bp

\* 160944 216116: contig of 55173 bp in length

\* 216117 229102: contig of 12886 bp in length.

\* 229102 229102: contig of 12886 bp in length.

\* 229102 229102: contig of 12886 bp in length.

\* 229102 229102: contig of 12886 bp in length.

\* 229102 229102: contig of 12886 bp in length.

\* 229102 229102: contig of 12886 bp in length.

\* 229102 229102: contig of 12886 bp in length.

\* 229102 229102: contig of 12886 bp in length.

\* 229102 229102: contig of 12886 bp in length.

\* 229102 229102: contig of 12886 bp in length.

\* 229102 229102: contig of 12886 bp in length.

\* 229102 229102: contig of 12886 bp in length.

\* 229102 229102: contig of 12886 bp in length.

\* 229102 229102: contig of 12886 bp in length.

\* 229102 229102: contig of 12886 bp in length.

\* 229102 229102: contig of 12886 bp in length.

\* 229102 229102: contig of 12886 bp in length.

\* 229102 229102: contig of 12886 bp in length.

\* 229102 229102: contig of 12886 bp in length.

\* 229102 229102: contig of 12886 bp in length.

\* 229102 229102: contig of 12886 bp in length.

\* 229102 229102: contig of 12886 bp in length.

\* 229102 229102: contig of 12886 bp in length.

\* 229102 229102: contig of 12886 bp in length.

\* 229102 229102: contig of 12886 bp in length.

\* 229102 229102: contig of 12886 bp in length.

\* 229102 229102: contig of 12886 bp in length.

\* 229102 229102: contig of 12886 bp in length.

\* 229102 229102: contig of 12886 bp in length.

\* 229102 229102: contig of 12886 bp in length.

\* 229102 229102: contig of 12886 bp in length.

\* 229102 229102: contig of 12886 bp in length.

\* 229102 229102: contig of 12886 bp in length.

\* 229102 229102: contig of 12886 bp in length.

\* 229102 229102: contig of 12886 bp in length.

\* 229102 229102: contig of 12886 bp in length.

misc\_feature 98241..120308  
/note="assembly\_fragment"  
misc\_feature 120409..160843  
/note="assembly\_fragment"  
misc\_feature 160944..216116  
/note="assembly\_fragment"  
misc\_feature 216217..229102  
/note="assembly\_fragment"  
clone\_end:T7  
vector\_side:right"

ORIGIN

Query Match 14.0%; Score 33.4; DB 2; Length 229102;  
Best Local Similarity 54.5%; Pred. No. 44;  
Matches 67; Conservative 0; Mismatches 56; Indels 0; Gaps 0;  
QY 50 GCCAACGGGAGTCGTGTTTCGGTTTCGATACAGATAAAACCCAGTCGATAGCCCTCGA 109  
Db 122376 GCGCCAGGGGGCTGTGATTTCTTTAGTTTCAGAGCACTCAGACTGCCAAATCCCTGTA 122435  
QY 110 CCGTCGCGTATATCTTAGAAGCAACAGCTAAAGNACTATTTCAAGAACTGTGTGGC 169  
Db 122436 CTCTGAGTTGTATTTTCTGGGCCCAAAAAAAGGGCTGAGTGGC 122495  
QY 170 AAG 172  
Db 122496 CAG 122498

Search completed: September 18, 2004, 12:08:37  
Job time : 1148.55 secs

This Page Blank (uspio)



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 2004, 09:40:06 ; Search time 257.309 Seconds  
(without alignments)  
3945.922 Million cell updates/sec

Title: US-10-614-282-1

Perfect score: 239

Sequence: 1 atcagtcacgacttgtaag.....tgtgttcgcggcggacaat 239

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N Geneseq 29Jan04:\*

1: geneseqn1980s:\*

2: geneseqn1990s:\*

3: geneseqn2000s:\*

4: geneseqn2001as:\*

5: geneseqn2001bs:\*

6: geneseqn2002s:\*

7: geneseqn2003as:\*

8: geneseqn2003bs:\*

9: geneseqn2003cs:\*

10: geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	239	100.0	18920	4	ABL02306
2	32.8	13.7	727	6	ABQ70202 Drosophil
3	32	13.4	3250	9	ABQ70202 Listeria
4	31.8	13.3	17468	6	ADB68937 C. neofor
5	31.6	13.2	969	7	ABA01441 Streptoco
6	31.5	13.2	110000	6	ACA36437 Prokaryot
7	31.2	13.1	11638	4	Continuation (23 o
8	31	13.0	536	6	ABL04264 Drosophil
9	31	13.0	536	6	ABQ30441 Oligonuc
10	31	13.0	6061	4	ABQ30440 Oligonuc
11	30.8	12.9	391	2	AAS45335 Chemical
12	30.8	12.9	1625	4	AAG60282 Human bra
13	30.8	12.9	1738	4	ABL20185 Drosophil
14	30.8	12.9	106286	6	ABL14353 Drosophil
15	30.8	12.9	110000	6	Continuation (5 of
16	30.8	12.9	341511	6	Continuation (3 of
17	30.2	12.6	1360	5	ABQ55200 Genomic D
18	30.2	12.6	1360	5	ABA21393 Human ner
19	30.2	12.6	2000	7	ADA71938 Human ner
20	30.2	12.6	167343	6	ABL64403 Rice gene
21	30.2	12.6	167343	6	ABL64403 Stomach c
22	30	12.6	873	6	ABL67239 Thyroid c
23	30	12.6	2377	6	ABQ68462 Listeria
					ABQ70841 Listeria

24	29.8	12.5	438	5	ABA20697 Human ner	
25	29.8	12.5	349980	5	AHA41224 Pyrococcu	
26	29.6	12.4	591	7	ACF73872 Staphyloc	
27	29.6	12.4	1225	2	AAG73396 CviJ1 ORF	
28	29.6	12.4	1472	2	AAV74866 Staphyloc	
C	28	29.6	12.4	5496	2	AAQ73395 CviJ1 cod
29	29.6	12.4	110000	6	Continuation (22 o	
30	29.6	12.4	6594	3	AAA03041_21	
31	29.4	12.3	6594	3	AAQ70287 Plasmodiu	
C	32	29.4	12.3	43064	7	ACC80906 Human CYP
33	29.2	12.2	468	3	AAC35767 Zea mays	
C	34	29.2	12.2	1254	6	ABK35684 cDNA sequ
C	35	29.2	12.2	6200	5	AAS72544 DNA encod
36	29	12.1	412	7	ABX46645 Bovine ES	
C	37	29	12.1	439	5	ABV15324 Human pro
C	38	29	12.1	3863	7	ACA23730 Prokaryot
39	28.8	12.1	777	7	ACA30362 Prokaryot	
40	28.8	12.1	1074	7	ACA52220 Prokaryot	
41	28.8	12.1	349901	9	ADC86940 Human GPC	
42	28.8	12.1	349938	9	ADC87621 Human GPC	
C	43	28.6	12.0	409	4	AAS37706 Novel hum
44	28.6	12.0	768	4	AAS52955 S. epider	
C	45	28.6	12.0	923	7	ABZ51987 Aspergill

## ALIGNMENTS

RESULT 1  
ABL02306  
ID ABL02306 standard; cDNA; 18920 BP.  
XX  
AC ABL02306;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 1400.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
XX pharmaceutical; gene; ss.  
XX  
OS Drosophila melanogaster.  
XX  
PN W0200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US009231.  
XX  
PR 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
XX  
XX WPI; 2001-656860/75.  
XX  
XX P-PSDB; ABB56203.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
XX genes from Drosophila and for elucidating cell signaling and cell-cell  
XX interactions.  
XX  
XX Claim 1; SEQ ID NO 1400; 2lpp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
XX ABB72072). The sequence data for this patent did not form part of the  
XX printed specification, but was obtained in electronic format directly



```

Db      335 CTTCCTCTCGATCTCTACAGACCGAAGGAGGAGCACTCCACCATAGCGATCTTTCCC 276
QY      127 TAGAAAGCAAAACAGCT 142
Db      275 TATACACATATCCCT 260

RESULT 4
ABA01441/c
ID ABA01441 standard; DNA; 17468 BP.
XX
AC ABA01441;
XX
DT 21-FEB-2002 (first entry)
XX
XX Streptococcus thermophilus eps3 operon #2.
DE
XX Exopolysaccharide; lactic acid bacterium; eps; fermented food product;
KW ds.
XX Streptococcus thermophilus.
OS
XX WO200179500-A2.
PN
XX 25-OCT-2001.
XX
PF 18-APR-2001; 2001WO-FR001199.
XX
PR 18-APR-2000; 2000FR-00004972.
XX
PA (INRG ) INRA INST NAT RECH AGRONOMIQUE.
PA (DANO-) CTE DANONE SA GERAVALS.
PA (RHOD ) RHODIA CHIM.
XX
PI Rallu F, Besancon-Yoshpe I, Fremaux C, Mengaud J, Renault P;
XX WPI; 2002-017616/02.
XX
XX New nucleic acid fragments containing exopolysaccharide operon, useful
PT e.g. for increasing exopolysaccharide synthesis in lactic acid bacteria.
XX
PS Claim 9; Page 75-80; 144pp; French.
XX
CC The present sequence is an eps operon from Streptococcus thermophilus.
CC Proteins encoded by the eps operon function in exopolysaccharide (EPS)
CC synthesis. The operon is useful for producing chimeric eps operons, for
CC optimising production of EPS in lactic acid bacteria. EPS impart texture,
CC mouth feel and rheological properties to fermented food products (e.g.
CC yoghurt). They function as thickeners, to provide free-flowing and creamy
CC texture, and may also have biological activities beneficial to health
XX
SQ Sequence 17468 BP; 5871 A; 2455 C; 3307 G; 5835 T; 0 U; 0 Other;
Query Match 13.3%; Score 31.8; DB 6; Length 17468;
Best Local Similarity 56.1%; Pred. No. 7.9;
Matches 60; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 85 TAAACCCAGCTCGATAGCCCTCGACCGTCGCGTAATATCTTTAGAAAGCAAAACAGCTAA 144
Db 7627 TAAAAAGCACGACAAATAGTCTTCAACAATATCTTTGATATCGGACCAATCGACCGATTAT 7568
QY 145 AGNACTATTTCAAGACTGTGTGGCAAGTGAAGGGTAGTTAGTATA 191
Db 7567 CGATATATTCACATCGAGTCTTACCACTCATCAAAAGTTGGTATA 7521

RESULT 5
ACA36437
ID ACA36437 standard; DNA; 969 BP.
XX
AC ACA36437;
XX

```

```

DT 19-JUN-2003 (first entry)
XX
DE Prokaryotic essential gene #18094.
XX
KW Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX
OS Listeria monocytogenes.
XX
XX WO200277183-A2.
PN
XX 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
XX 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
PA
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR P-ESDB; ABU32567.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 14; SEQ ID NO 24307; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 969 BP; 320 A; 166 C; 190 G; 293 T; 0 U; 0 Other;
Query Match 13.2%; Score 31.6; DB 7; Length 969;
Best Local Similarity 65.7%; Pred. No. 2.7;
Matches 46; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

```



```

PF 01-SEP-2001; 2001WO-EP010074.
XX
XX
PR 01-SEP-2000; 2000DE-01043826.
PR 05-SEP-2000; 2000DE-01044543.
XX
XX
PA (EPiG-) EPIGENOMICS AG.
XX
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
XX
DR WPI; 2002-371829/40.
XX
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful for
PT diagnosis and prognosis, comprises selective hybridization of amplicons
PT from chemically treated DNA.
XX
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one member,
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly by detecting mutations or single nucleotide polymorphisms
CC (SNP's); and (ii) for differentiation of cell or tissue types and for
CC investigating cell differentiation. The method allows the methylation
CC status of many C residues to be determined simultaneously. ABQ13410-
CC ABQ54121 represent genomic DNA sequences used to illustrate the method
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention
XX
XX
SQ Sequence 536 BP; 189 A; 188 C; 65 G; 94 T; 0 U; 0 Other;

Query Match 13.0%; Score 31; DB 6; Length 536;
Best Local Similarity 64.8%; Pred. No. 3.4;
Matches 46; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

Qy 34 CGTCGTCGTCATCGCCACCGGAGTCGTTTCGTTTCGATACAGATAAAACCCA 93
Db 193 CGTCGTCGTCGTCATCGCCACCGGAGTCGTTTCGTTTCGATACAGATAAAACCCA 93
Qy 94 CGTCGATAGCC 104
Db 253 CGCATCCGCC 263

RESULT 9
ABQ30440/c
ID ABQ30440 standard; DNA; 536 BP.
XX
XX
AC ABQ30440;
XX
XX
XX 12-JUL-2002 (first entry)
XX
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 17031.
XX
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
XX
OS Homo sapiens.
XX
XX
PN WO200218632-A2.
XX
XX
PD 07-MAR-2002.

```

```

XX
XX
PF 01-SEP-2001; 2001WO-EP010074.
XX
XX
PR 01-SEP-2000; 2000DE-01043826.
PR 05-SEP-2000; 2000DE-01044543.
XX
XX
PA (EPiG-) EPIGENOMICS AG.
XX
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
XX
DR WPI; 2002-371829/40.
XX
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful for
PT diagnosis and prognosis, comprises selective hybridization of amplicons
PT from chemically treated DNA.
XX
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one member,
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly by detecting mutations or single nucleotide polymorphisms
CC (SNP's); and (ii) for differentiation of cell or tissue types and for
CC investigating cell differentiation. The method allows the methylation
CC status of many C residues to be determined simultaneously. ABQ13410-
CC ABQ54121 represent genomic DNA sequences used to illustrate the method
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention
XX
XX
SQ Sequence 536 BP; 94 A; 65 C; 188 G; 189 T; 0 U; 0 Other;

Query Match 13.0%; Score 31; DB 6; Length 536;
Best Local Similarity 64.8%; Pred. No. 3.4;
Matches 46; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

Qy 34 CGTCGTCGTCATCGCCACCGGAGTCGTTTCGTTTCGATACAGATAAAACCCA 93
Db 344 CGTCGTCGTCATCGCCACCGGAGTCGTTTCGTTTCGATACAGATAAAATCCAC 285
Qy 94 CGTCGATAGCC 104
Db 284 CGCATCCGCC 274

RESULT 10
AAS45335/c
ID AAS45335 standard; DNA; 6061 BP.
XX
XX
AC AAS45335;
XX
XX
XX 18-DEC-2001 (first entry)
XX
XX
DE Chemically pretreated complementary DNA associated with cell cycle #20.
XX
XX
KW Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;
KW human immunodeficiency virus; neurodegenerative disorder; solid tumour;
KW graft-versus-host disease; glomerular disease; Lewy body disease; cancer;
KW arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;
KW immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;
XX
XX
XX PCR primer.
XX
XX
OS Homo sapiens.
XX

```

```
PN WO200168911-A2.
XX
XX 20-SEP-2001.
XX
XX 15-MAR-2001; 2001WO-BP002945.
XX
XX 15-MAR-2000; 2000DE-01013847.
XX
XX 06-APR-2000; 2000DE-01019058.
XX
XX 07-APR-2000; 2000DE-01019173.
XX
XX 30-JUN-2000; 2000DE-01032529.
XX
XX 01-SEP-2000; 2000DE-01043826.
XX
XX (EPiG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2001-602751/68.
XX
XX Designing primers and probes for analyzing diseases associated with
XX cytosine methylation state e.g. arthritis, cancer, aging
XX arteriosclerosis comprising fragments of chemically modified genes
XX associated with cell cycle.
XX
XX Claim 1; SEQ ID NO 40; 28pp; English.
XX
XX Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA
XX molecules associated with the cell cycle and specific PCR primers of the
XX invention. The sequences are useful for detecting the methylation state
XX of all CpG dinucleotides in a sequence and therefore for analysing
XX associated diseases. By analysing cytosine methylations in the pretreated
XX DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
XX of existing diseases or the predisposition to specific diseases can be
XX ascertained. The parameters may be compared to another set of genetic
XX and/or epigenetic parameters, the differences serving as basis for
XX diagnosis and/or prognosis events which are disadvantageous to patients.
XX The sequences of the invention are useful for the diagnosis and therapy
XX of HIV infection, neurodegenerative disorders, graft-versus-host disease,
XX aging, glomerular disease, Lewy body disease, arthritis,
XX arteriosclerosis, solid tumours and cancers
XX
XX Sequence 6061 BP; 1558 A; 264 C; 1557 G; 2682 T; 0 U; 0 Other;
SQ
Query Match 13.0%; Score 31; DB 4; Length 6061;
Best Local Similarity 64.8%; Pred. No. 9.6;
Matches 46; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
QY 34 CGTCGTCGTCATCCCAACGGAGTCGTGTTTTCGGTTCGATACAGATAAACCCCA 93
DB 2164 CGTCGTCGCGGCCATCTCGACGCGACTCGCTTCTTTTCGATTCTACCTAATAAATCCAC 2105
QY 94 CGTCGATAGCC 104
DB 2104 CGCCATCCGCC 2094
RESULT 11
AAQ60282
ID AAQ60282 standard; DNA; 391 BP.
XX
XX AAQ60282;
XX
XX 25-MAR-2003 (revised)
XX 16-MAR-1994 (first entry)
XX
XX Human brain Expressed Sequence Tag EST02279.
XX
XX Gene transcription product; genetic markers; tagging; in vivo;
XX transcription; mapping; locations; chromosomes; chromosomal; ss.
XX
XX Homo sapiens.
XX
XX WO9316178-A2.
XX
```

```
PD 19-AUG-1993.
XX
XX 12-FEB-1993; 93WO-US001294.
XX
XX 12-FEB-1992; 92US-00837195.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICE.
XX
XX Venter CJ, Adams MD, Moreno RF;
XX
XX WPI; 1993-272882/34.
XX
XX Enriched oligonucleotides and corresp. sequences - used as markers for
XX human genes transcribed in-vivo, facilitate tagging of most human genes.
XX
XX Example 4; Page 310; 500pp; English.
XX
XX The Expressed Sequence Tag was isolated from a human brain cDNA library
XX as part of a large set of ESTs which can be used as markers for human
XX genes transcribed in vivo. They can be used to facilitate tagging of most
XX human genes, for mapping locations of expressed genes on chromosomes, for
XX individual or forensic identification, for mapping locations of disease-
XX associated genes, for identification of tissue type, and for prepn. of
XX antisense sequences, probes and constructs. EST02279 has a "poor" coding
XX probability as evaluated using the coding-region prediction program CRM.
XX See also AAQ59041-Q61440. (Updated on 25-MAR-2003 to correct PN field.)
XX
XX Sequence 391 BP; 96 A; 94 C; 89 G; 109 T; 0 U; 3 Other;
SQ
Query Match 12.9%; Score 30.8; DB 2; Length 391;
Best Local Similarity 54.6%; Pred. No. 3.5;
Matches 59; Conservative 0; Mismatches 49; Indels 0; Gaps 0;
QY 117 GTAATATCTTAGAAGCAACAGCTAAAGAACTATTTCAAGAACTGTGCGCAAGTGA 176
DB 33 GCAGCATCCTTACCTTGCAGAGTACTGATCTCGCTTCATGATTTTATGTCAAGTAAA 92
QY 177 GGGTAGTTAGTCATACACCGGTTATATCGAGTGGCGAGAAAGTGTGG 224
DB 93 TGGGTTTATGTCATCCCTAGTTCATGTCATGTCGAGAAAAGGGG 140
RESULT 12
ABL20185
ID ABL20185 standard; DNA; 1625 BP.
XX
XX ABL20185;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 12028.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ds.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
XX
XX 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers BW;
XX
XX WPI; 2001-656860/75.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX
```



Db 10305 TAGGATTGAAAAAATTATTAT 10280

Search completed: September 18, 2004, 11:29:54  
Job time : 261.309 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 2004, 11:14:16 ; Search time 59.8737 Seconds  
(without alignments)  
2215.217 Million cell updates/sec

Title: US-10-614-282-1  
Perfect score: 239  
Sequence: 1 atcagtacactgtgtaag.....tgtggtccggctggacaat 239

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29.6	12.4	1225	1	US-08-181-629A-3
2	29.6	12.4	1472	4	US-08-956-171E-555
3	29.6	12.4	5496	1	US-08-181-629A-2
4	28.6	12.0	1731	4	US-09-134-001C-1118
5	27.8	11.6	423	4	US-09-397-787-196
6	27.6	11.5	716	3	US-08-998-416-746
7	27.4	11.5	1296	4	US-09-543-681A-2060
8	27.4	11.5	1650	4	US-09-107-532A-3186
9	27	11.3	1200	4	US-09-358-082A-27
10	27	11.3	1521	4	US-09-328-352-1397
11	27	11.3	2026	2	US-08-993-228-3
12	27	11.3	15071	4	US-09-358-082A-29
13	26.8	11.2	505	4	US-09-621-976-15639
14	26.8	11.2	612	4	US-09-716-865-7
15	26.8	11.2	1975	4	US-09-910-174B-3
16	26.8	11.2	1975	4	US-09-620-461-3
17	26.8	11.2	2229	4	US-09-910-174B-1
18	26.8	11.2	2229	4	US-09-620-461-1
19	26.8	11.2	10754	2	US-08-966-958-1
20	26.8	11.2	10754	2	US-09-215-817-1
21	26.8	11.2	10754	3	US-09-342-353-1
22	26.8	11.2	53216	4	US-09-716-865-23
23	26.6	11.1	657	4	US-09-489-039A-2668
24	26.6	11.1	1128	4	US-09-489-039A-4389
25	26.6	11.1	2147	1	US-08-448-600-1
26	26.6	11.1	3088	3	US-08-937-610-3
27	26.6	11.1	4047	4	US-09-534-407-4

28	26.6	11.1	4047	4	US-09-999-201B-3
29	26.6	11.1	9610	4	US-09-566-921-45
30	26.6	11.1	640681	4	US-09-790-988-1
31	26.4	11.0	2515	3	US-08-714-918-37
32	26.4	11.0	2515	3	US-09-265-315-37
33	26.4	11.0	2515	3	US-09-265-315-37
34	26.4	11.0	2515	3	US-09-266-417-37
35	26.4	11.0	2515	4	US-09-528-709-37
36	26.4	11.0	2515	4	US-09-527-745-37
37	26.4	11.0	7762	4	US-08-956-171E-124
38	26.4	11.0	1230025	4	US-09-198-452A-1
39	26.2	11.0	87350	3	US-08-781-891-79
40	26.2	11.0	87350	4	US-09-618-166-79
41	26.2	11.0	87543	4	US-09-791-211-3
42	26	10.9	956	4	US-09-641-638-42
43	26	10.9	956	4	US-09-641-638-43
44	26	10.9	1001	4	US-09-641-638-620
45	26	10.9	1001	4	US-09-641-638-621

#### ALIGNMENTS

RESULT 1  
US-08-181-629A-3  
; Sequence 3, Application US/08181629A  
; Patent No. 5472872  
; GENERAL INFORMATION:  
; APPLICANT: Swaminathan, Neela  
; APPLICANT: Van Etten, James  
; APPLICANT: Mead, David  
; APPLICANT: Skowron, Piotr  
; TITLE OF INVENTION: Recombinant CviJI Restriction Endonuclease  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; FILING DATE:  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Clough, David W.  
; REGISTRATION NUMBER: 36,107  
; REFERENCE/DOCKET NUMBER: 31504  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/474-6300  
; TELEFAX: 312/474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1225 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: join(1..33, 55..1128)  
US-08-181-629A-3

Query Match 12.4%; Score 29.6; DB 1; Length 1225;  
Best Local Similarity 64.7%; Pred. No. 1.1;  
Matches 44; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 131 AAGCAACAGCTAAAGAACTATTTCAGAACTGTGTGGCAAGTGAAGGCTAGTTAGTAT 190  
Db 691 AAGTTCACAGGAAAGAAATAGAGAGTTCTATCGTTCAGAGAAAGTAGTTAGTATG 750

QY 191 ACACCGGT 198  
Db 751 GCACCGGT 758

## RESULT 2

US-08-956-171E-555/c  
; Sequence 555, Application US/08956171E  
; Patent No. 6593114

## GENERAL INFORMATION:

APPLICANT: Charles Kunsch  
; Gil H. Choi  
; Patrick S. Dillon  
; Craig A. Rosen  
; Steven C. Barash  
; Michael R. Fannon

TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences

NUMBER OF SEQUENCES: 5256

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/956,171E

FILING DATE: 20-Oct-1997

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/009,861

FILING DATE: January 5, 1996

APPLICATION NUMBER: 08/781,986

FILING DATE: January 3, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Mark J. Hyman

REGISTRATION NUMBER: 46,789

REFERENCE/DOCKET NUMBER: PB248P1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (240) 314-1224

TELEFAX: (301) 309-8439

INFORMATION FOR SEQ ID NO: 555:

SEQUENCE CHARACTERISTICS:

LENGTH: 1472 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 555:

US-08-956-171E-555

Query Match 12.4%; Score 29.6; DB 4; Length 1472;  
Best Local Similarity 50.0%; Pred. No. 1.2;  
Matches 74; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 60 GTCGTGTTTCGGTTCGATACAGATAAAACCCAGTCGATAGCCCTGACCGTCGGTA 119

Db 1016 GTATCGTGCTTATTGTGTGGAATATCAITCAACGGTATCATCTGTACACGGACATA 957

QY 120 ATATTCTTAAAGCAACAGCTAAAGAACTATTTCAGAACTGTGTGCAAGTGAAGG 179

Db 956 CTACACAGTAGTACCCATACCAATGAGGTATTTCAGGACTGCTCAAAATTTCAATG 897

QY 180 TAGTGTAGTATACACCGGTTATATCGGA 207

Db 896 TAATCGGTATCATTCATTCAATTAATGCTGA 869

## RESULT 3

US-08-181-629A-2

; Sequence 2, Application US/08181629A

; Patent No. 5472872

GENERAL INFORMATION:

APPLICANT: Swaminathan, Neela

; APPLICANT: Van Ebben, James

APPLICANT: Mead, David

APPLICANT: Skowron, Piotr

TITLE OF INVENTION: Recombinant CviJI Restriction Endonuclease

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun

; STREET: 6300 Sears Tower, 233 South Wacker Drive

; CITY: Chicago

; STATE: Illinois

; COUNTRY: United States of America

; ZIP: 60606-6402

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/181,629A

FILING DATE:

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Clough, David W.

REGISTRATION NUMBER: 36,107

REFERENCE/DOCKET NUMBER: 31504

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312/474-6300

TELEFAX: 312/474-0448

TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 5496 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-181-629A-2

Query Match 12.4%; Score 29.6; DB 1; Length 5496;

Best Local Similarity 64.7%; Pred. No. 2.2;

Matches 44; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 131 AAGCAACAGCTAAAGAACTATTTCAGAACTGTGTGCAAGTGAAGGCTAGTTAGTAT 190

Db 1759 AAGTTCACAGGAAAGAAATAGAGAGTTCTATCGTTCAGAGAAAGTAGTTAGTATG 1818

QY 191 ACACCGGT 198

Db 1819 GCACCGGT 1826

## RESULT 4

US-09-134-001C-1118

; Sequence 1118, Application US/09134001C

; Patent No. 6380370

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: GTC-007

CURRENT APPLICATION NUMBER: US/09/134,001C

CURRENT FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/064,964

```
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,773
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1118
; LENGTH: 1731
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1118

Query Match      12.0%; Score 28.6; DB 4; Length 1731;
Best Local Similarity 59.0%; Pred. No. 2.9;
Matches 49; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

Qy 77 GATACAGATAAAACCCAGCTGATAGCCTCGACGGTCGCTATATCTTAGAAGCAA 136
Db 1618 GAAACACATCAACGCGAAAAACAAAGCACACACAGCTCAACAAATAGATATTGCTCAA 1677

Qy 137 ACAGCTAAAGAACTATTTCAGA 159
Db 1678 AAAGCTAAAGACTATTGGTGA 1700

RESULT 5
US-09-397-787-196
; Sequence 196, Application US/09397787
; Patent No. 6468758
; GENERAL INFORMATION:
; APPLICANT: Benson, Darin R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR OVARIAN
; TITLE OF INVENTION: CANCER THERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.466C2
; CURRENT APPLICATION NUMBER: US/09/397,787
; CURRENT FILING DATE: 1999-09-16
; NUMBER OF SEQ ID NOS: 334
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 196
; LENGTH: 423
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-397-787-196

Query Match      11.6%; Score 27.8; DB 4; Length 423;
Best Local Similarity 57.5%; Pred. No. 2.8;
Matches 50; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

Qy 99 ATAGCCCTCGACCGTCGCTAATATTCTTAGAAGCAACAGCTAAAGCACTATTTCAG 158
Db 66 ATTCCCTCCCTCCCTAGCAAAATATTATTAAAAAAGCAACAGGCGGAG 125

Qy 159 AACTGTGTGGCAAGTGAAGGTAGTTA 185
Db 126 AGGAGGATAGGAGGGACATTGATA 152

RESULT 6
US-08-998-416-746
; Sequence 746, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Rebeschung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHEYA GOSSYPII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
```

```
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwalls Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 746:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 716 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAG1484UP
US-08-998-416-746

Query Match      11.5%; Score 27.6; DB 3; Length 716;
Best Local Similarity 55.1%; Pred. No. 4.3;
Matches 54; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

Qy 16 GTAAGCGCGCAGGCGACGCTGCTGCTCATCGCCAAACGGGAGTCGTGTTTCGTT 75
Db 454 GTACGAGCCGATGTTGAACGACTACGTGCTCAAGTCGAGACGAACTTTGATACAGTA 513

Qy 76 CGATACAGATAAAACCCACGCTCGATAGCCCTCGACCGT 113
Db 514 CGAGCTATACAAAAGGCGCTGGAGCACAATCGGCAGT 551

RESULT 7
US-09-543-681A-2060/C
; Sequence 2060, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 2060
; LENGTH: 1296
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-2060

Query Match      11.5%; Score 27.4; DB 4; Length 1296;
Best Local Similarity 55.9%; Pred. No. 6.8;
Matches 52; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
```

QY 66 TTTTTCGGTTCATACAGATAAAACCCAGCTCGATAGCCCTCGACCGTCGCGTAATATTC 125  
DB 1235 TTGCTATGAGCATACATGAAGCGCTGCTTCAAAAGCCGACGCGAGGTAATACGT 1176  
QY 126 TTAGAAAGCAACAGCTAAAGAACTATTTCAG 158  
DB 1175 TTTTCTAACATTAAGTGGAGACTGTTTAAG 1143

RESULT 8  
US-09-107-532A-3186/c  
; Sequence 3186, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESSER: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD/ROM ISO9660  
; COMPUTER: PC  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,532A  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/085,598  
; FILING DATE: 14 May 1998  
; APPLICATION NUMBER: 60/051571  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ariniello, Pamela Deneka  
; REGISTRATION NUMBER: 40,489  
; REFERENCE/DOCKET NUMBER: GTC-012  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (781)893-5007  
; TELEFAX: (781)893-8277  
; INFORMATION FOR SEQ ID NO: 3186:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1650 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: circular  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Enterococcus faecium  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (B) LOCATION 1...1650  
; SEQUENCE DESCRIPTION: SEQ ID NO: 3186:  
US-09-107-532A-3186  
Query Match 11.5%; Score 27.4; DB 4; Length 1650;  
Best Local Similarity 59.7%; Pred. No. 7.6;  
Matches 46; Conservative 0; Mismatches 31; Indels 0; Gaps 0;  
QY 33 ACGTCGTCGTCGATCGCAACCGGAGTCGTGTTTTTCGGTTTCGATACAGATAAAACCC 92  
DB 206 ACGTCATACCATTCATCGCAATGAGATGATGATCGGATTCGATTCAGATAAGCCG 147  
QY 93 ACGTCGATAGCCCTCGA 109

Db 146 ATACCAAGATCTTTCAA 130

## RESULT 9

US-09-358-082A-27  
; Sequence 27, Application US/09358082A  
; Patent No. 6589784  
; GENERAL INFORMATION:  
; APPLICANT: Antoniou, Michael  
; APPLICANT: Crombie, Robert  
; TITLE OF INVENTION: A Polynucleotide  
; CURRENT APPLICATION NUMBER: US/09/358,082A  
; CURRENT FILING DATE: 1999-07-21  
; PRIOR APPLICATION NUMBER: GB 9815879.3  
; PRIOR FILING DATE: 1998-07-21  
; PRIOR APPLICATION NUMBER: US 60/107688  
; PRIOR FILING DATE: 1998-11-09  
; PRIOR APPLICATION NUMBER: GB 9906712.6  
; PRIOR FILING DATE: 1999-03-23  
; PRIOR APPLICATION NUMBER: US 60/127410  
; PRIOR FILING DATE: 1999-04-01  
; PRIOR APPLICATION NUMBER: GB 9909494.8  
; PRIOR FILING DATE: 1999-04-23  
; PRIOR APPLICATION NUMBER: US 60/134016  
; PRIOR FILING DATE: 1999-05-12  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 27  
; LENGTH: 1200  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-358-082A-27

Query Match 11.3%; Score 27; DB 4; Length 1200;  
Best Local Similarity 54.5%; Pred. No. 9.1;  
Matches 54; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 71 CGTTTCGATACAGATAAAACCCAGCTCGATAGCCCTCGACCGTCGCGTAATATTCCTAGA 130  
DB 362 CAGCTAAATACAGTTGCAACCCACATTCGAATGATATACATTTGGAATGAGGTCCTTGA 421

QY 131 AAGCAACACAGCTAAAGAACTATTTCAGAACTGTGTGGC 169  
DB 422 AGTTAAATCCITTAACCATGATTCACACCATTTGCTTAGC 460

## RESULT 10

US-09-328-352-1397  
; Sequence 1397, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 1397  
; LENGTH: 1521  
; TYPE: DNA  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-1397

Query Match 11.3%; Score 27; DB 4; Length 1521;  
Best Local Similarity 60.0%; Pred. No. 10;  
Matches 45; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 115 GCGTAATATTCCTAGAAACCAACAGCTAAAGAACTATTTCAGAACTGTGTGGCAAGTG 174  
DB 1296 GTGGAATATGCTTAAAGAGCGAAATGCTATAGAACAAATTTGGTCAACATTTTGGCCATGA 1355

```
Qy 175 AAGGTAGTGTAGTA 189
Db 1356 TTGTTTGTAGTGTGA 1370

RESULT 11
US-08-993-228-3
; Sequence 3, Application US/08993228
; Patent No. 5976838
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: LaVallie, Edward R.
; APPLICANT: Racie, Lisa A.
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; APPLICANT: Agostino, Michael J.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/993.228
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sprunger, Suzanne A.
; REGISTRATION NUMBER: 41,323
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8284
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2026 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-993-228-3

Query Match 11.3%; Score 27; DB 2; Length 2026;
Best Local Similarity 53.3%; Pred. No. 12;
Matches 57; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Qy 63 GTGTTTTCGTTTCGATACAGATAAAACCCACGTCGATAGCCCTCGACCGTCGCGTAATA 122
Db 1181 GTGCTTTCCTACTTCCTAAAGACCCGACGCTGCTGGGAACCATGTGTTCTCTGAGTA 1240

Qy 123 TTCTTAGAAGCAACAGCTAAAGAACTATTTCAGAACTGTGTGGC 169
Db 1241 TTCTCAGAGGTTAAACAAGTCACAGTCAGCTTCTGAAATTAGTGTC 1287

RESULT 12
US-09-358-082A-29
; Sequence 29, Application US/09358082A
; Patent No. 6589784
; GENERAL INFORMATION:
; APPLICANT: Antoniou, Michael
; APPLICANT: Crombie, Robert
; TITLE OF INVENTION: A Polynucleotide

FILE REFERENCE: CAC00056
CURRENT APPLICATION NUMBER: US/09/358.082A
CURRENT FILING DATE: 1999-07-21
PRIOR APPLICATION NUMBER: GB 9815879.3
PRIOR FILING DATE: 1998-07-21
PRIOR APPLICATION NUMBER: US 60/107688
PRIOR FILING DATE: 1998-11-09
PRIOR APPLICATION NUMBER: GB 9906712.6
PRIOR FILING DATE: 1999-03-23
PRIOR APPLICATION NUMBER: US 60/127410
PRIOR FILING DATE: 1999-04-01
PRIOR APPLICATION NUMBER: GB 9909494.8
PRIOR FILING DATE: 1999-04-23
PRIOR APPLICATION NUMBER: US 60/134016
PRIOR FILING DATE: 1999-05-12
NUMBER OF SEQ ID NOS: 29
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 29
LENGTH: 15071
TYPE: DNA
ORGANISM: Homo sapiens
US-09-358-082A-29

Query Match 11.3%; Score 27; DB 4; Length 15071;
Best Local Similarity 54.5%; Pred. No. 32;
Matches 54; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Qy 71 CGTTTCGATACAGATAAAACCCACGTCGATAGCCCTCGACCGTCGTAATATCTTAGA 130
Db 13843 CAGCTAAATACAGTTGCAACCCACATTCGAAGTGGTATACATTGGAATGAGGTCCTTGA 13902

Qy 131 AAGCAACAGCTAAAGAACTATTTCAGAACTGTGTGGC 169
Db 13903 AGTTAAATCTCTAAACCATGATTCAAAACCATTCGTTAGC 13941

RESULT 13
US-09-621-976-15639
; Sequence 15639, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621.976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 15639
; LENGTH: 505
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-15639

Query Match 11.2%; Score 26.8; DB 4; Length 505;
Best Local Similarity 14.6%; Pred. No. 7;
Matches 19; Conservative 62; Mismatches 49; Indels 0; Gaps 0;

Qy 10 GACTTGTAGCGCGCAGGAGGAGCAGTCGTCGTCGTCATCGCCACGCGAGTCGTGTTT 69
Db 39 GKYTSGCGGCGSRGWCXKMGYKWSKMGKYSKMGKYSKMGKYSKMGKYSKMGKYSKMGKYS 98

Qy 70 TCGGTTTCGATACAGATAAAACCCACGTCGATAGCCCTCGACCGTCGCGTAATATCTTAG 129
Db 99 KCRKTYCKRGWWSKWRMWRMWRMWRMWRMWRMWRMWRMWRMWRMWRMWRMWRMWRMWRM 158

Qy 130 AAAGCAACA 139
Db 159 RMCKSSTRYM 168
```

**Qy** 121 TATTCTTAGAAAGCAACAGCTAAGAAGACTATTTCAAGAACTGTGGCAGTGAAGGTT 180  
||||| | | | | | | | | | | | | | | | |  
**Dβ** 553 TATTCTTCAAAGACACAACAAAAGACCTGTACCACAACAAAGAGGGAGTGTAACGT 612

Result No.	Score	Query Match	Length	DB	ID	Description
1	239	100.0	239	17	US-10-614-282-1	Sequence 1, Appli
C 2	32.8	13.7	727	16	US-10-398-221-3015	Sequence 3015, Ap
C 3	32.8	13.7	917	13	US-10-424-599-14250	Sequence 14250, A
C 4	32.8	13.7	2731748	17	US-10-297-465A-1	Sequence 1, Appli
C 5	32	13.4	3250	16	US-10-320-797-64	Sequence 64, Appl
C 6	31.6	13.2	969	13	US-10-282-122A-24307	Sequence 24307, A
C 7	31.4	13.1	661	17	US-10-767-701-1632	Sequence 1632, Ap
C 8	31.4	13.1	1205	13	US-10-390-184-2	Sequence 2, Appli
C 9	31.4	13.1	1207	17	US-10-437-963-35457	Sequence 35457, A
C 10	31.4	13.1	1238	13	US-10-390-184-3	Sequence 3, Appli
C 11	31.4	13.1	158405	15	US-10-375-223-86	Sequence 86, Appl
C 12	31	13.0	6061	15	US-10-239-676-40	Sequence 40, Appl
C 13	30.2	12.6	498	17	US-10-437-963-92822	Sequence 92822, A
C 14	30.2	12.6	616	13	US-10-027-632-203363	Sequence 203363,

[illegible]

OY 181 AGTTAGTGATACACCGGTTATATCGAGTGGCGAGAAAGTGTGGTTCGGCTGGACAAT 239  
|||||  
Db 181 AGTTAGTGATACACCGGTTATATCGAGTGGCGAGAAAGTGTGGTTCGGCTGGACAAT 239  
|||||

## RESULT 2

US-10-398-221-3015/c  
; Sequence 3015, Application US/10398221  
; Publication No. US20040018514A1  
; GENERAL INFORMATION:  
; APPLICANT: KINST, Frederik  
; APPLICANT: GLASER, Philippe  
; TITLE OF INVENTION: Listeria innocua, genome and applications  
; FILE REFERENCE: 344 702 - US  
; CURRENT APPLICATION NUMBER: US/10/398,221  
; CURRENT FILING DATE: 2003-03-27  
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061  
; PRIOR FILING DATE: 2001-10-04  
; PRIOR APPLICATION NUMBER: FR 00/12 697  
; PRIOR FILING DATE: 2000-10-04  
; NUMBER OF SEQ ID NOS: 4025  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 3015  
; LENGTH: 727  
; TYPE: DNA  
; ORGANISM: Listeria monocytogenes 4b  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(end)  
; OTHER INFORMATION: n can be any nucleotide: a,g,c or t/u

US-10-398-221-3015

Query Match 13.7%; Score 32.8; DB 16; Length 727;  
Best Local Similarity 64.5%; Pred. No. 1;  
Matches 49; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

OY 88 AACCCACGTCGTAGCCCTCGACCGTCGCGTAATATCTTAGAAGCAACAGCTAAAGA 147  
|||||  
Db 214 AACCAAAATTGATAGCCCTCGTCGTATATCTTACACCAAAAAGTAAACTA 155  
|||||

OY 148 ACTATTTCAAGAACTG 163  
|||||  
Db 154 GACAAACAGATCTG 139  
|||||

## RESULT 3

US-10-424-599-14250  
; Sequence 14250, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 14250  
; LENGTH: 917  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_112876C.1

US-10-424-599-14250

Query Match 13.7%; Score 32.8; DB 13; Length 917;  
Best Local Similarity 61.9%; Pred. No. 1.2;  
Matches 52; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

OY 124 TCTTAGAAGCAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGTAGT 183  
|||||  
Db 246 TCTTAGCAAACTACTACTACAGAAAGCTTCGAGAGAGTTTGCARAAATTTGGTGAAGT 305  
|||||  
OY 184 TAGTGATACACCGGTTATATCGGA 207  
|||||  
Db 306 TGTTCAATGCAAGGGTTGTAACCTGA 329  
|||||

## RESULT 4

US-10-297-465A-1/c  
; Sequence 1, Application US/10297465A  
; Publication No. US20040142413A1  
; GENERAL INFORMATION:  
; APPLICANT: Simpson, Andrew  
; APPLICANT: Reinach, Fernando  
; APPLICANT: Setubal, Joao  
; APPLICANT: Medanis, Joao  
; APPLICANT: Arruda, Paulo  
; TITLE OF INVENTION: Isolated Genome of Xylella Fastidiosa and Uses Thereof  
; FILE REFERENCE: FAPESP 202 US (10213376)  
; CURRENT APPLICATION NUMBER: US/10/297,465A  
; CURRENT FILING DATE: 2001-06-07  
; PRIOR APPLICATION NUMBER: PCT/IB01/01618  
; PRIOR FILING DATE: 2001-06-07  
; PRIOR APPLICATION NUMBER: 60/209,906  
; PRIOR FILING DATE: 2001-06-17  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 1  
; LENGTH: 2731748  
; TYPE: DNA  
; ORGANISM: Xylella fastidiosa

US-10-297-465A-1

Query Match 13.7%; Score 32.8; DB 17; Length 2731748;  
Best Local Similarity 53.0%; Pred. No. 61;  
Matches 70; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

OY 40 CCGTCGTATCGCAACGGAGTCGTGTTCCTCGATACAGATAAAACCCAGCTCGA 99  
|||||  
Db 547279 CCTCCAGATGGCAACGGCTGGCCGCTTCAATCATTTAAAAAGGCAATACAGACATTTGT 547220  
|||||

OY 100 TAGCCCTCGACCGTCGCGTAATATTTAGAAAGCAACAGCTAAAGAACTATTTCAAGA 159  
|||||

Db 547219 CATCCCTTAACAGTCGCGCTCATCATCAGCAACCCACCACCCGCTGTTTCAAGC 547160  
|||||

OY 160 ACTGTGTGSCAA 171  
|||||

Db 547159 AACCTGCACCAA 547148  
|||||

## RESULT 5

US-10-320-797-64/c  
; Sequence 64, Application US/10320797  
; Publication No. US20040014955A1  
; GENERAL INFORMATION:  
; APPLICANT: Eroshkin, Alexey M.  
; APPLICANT: Zamudio, Carlos  
; TITLE OF INVENTION: IDENTIFICATION OF ESSENTIAL GENES OF CRYPTOCOCCUS NEOFORMANS AND  
; TITLE OF INVENTION: METHODS OF USE  
; FILE REFERENCE: 10182-021-999  
; CURRENT APPLICATION NUMBER: US/10/320,797  
; CURRENT FILING DATE: 2002-12-16  
; PRIOR APPLICATION NUMBER: 60/341,261  
; PRIOR FILING DATE: 2001-12-17  
; NUMBER OF SEQ ID NOS: 3361  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 64  
; LENGTH: 3250  
; TYPE: DNA  
; ORGANISM: Cryptococcus neoformans

US-10-320-797-64



Db 7 GAACAAAGAGAAAAATTATTAGAAATTCATAATCTAAAGCAATACITTTTAAACAAGGTACT 66

Qy 168 GCAAGTGAAG 177

Db 67 GCAAGTGAAG 76

```

RESULT 7
US-10-767-701-1632/c
; Sequence 1632, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21 (53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 1632
; LENGTH: 661
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS27424_1
US-10-767-701-1632

```

Query Match	13.1%;	Score 31.4;	DB 13;	Length 1205;
Best Local Similarity	57.7%;	Pred. No. 4.1;		
Matches	56;	Conservative 0;	Mismatches 41;	Indels 0; Gaps 0;
QY	34	CGTCGTCGTCGTCATCGCCACGGGAGTCGTGTTTTTCGGTTCGATACAGATAAAACCCA	93	
Db	1035	CGTCGTCGTCGTCGTCATTAACGGGACACGTACGTCGTGTTTCGGGCCATATATATCTCG	1094	
OY	94	CGTCGATAGCCCTCGACCGTCGCGTAATATCTTAGA	130	



```

RESULT 14
US-10-027-632-203363
; Sequence 203363, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome

```

```

RESULT 15
US-10-027-632-203363
; Sequence 203363, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 203363
; LENGTH: 616
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-203363

Query Watch 12.6%; Score 30.2; DB 16; Length 616;
Best Local Similarity 56.6%; Pred. No. 7.8;

```

Matches	56;	Conservative	0;	Mismatches	43;	Indels	0;	Gaps	0;
QY	111	CGTCGGTAATATCTTTAGAAAGCAACAGCTAAAGAACTATTTCAAGAACTGTGTGGCA	170						
Db	275	CTTTGCAGAGGCTTCTTATCAAGCTAAGAGGGGCGAGAAATTCCTATTACTGTGTGTCATG	334						
QY	171	AGTGAAGGTAAGTCTAGTGTATACACCGTTATATCGGAGT	209						
Db	335	CATAAATGGTAATGCTGAACCTCCACTTACAGCATATT	373						

Search completed: September 18, 2004, 15:14:57  
 Job time : 269.741 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 2004, 10:33:16 ; Search time 2018.88 Seconds  
(without alignment)  
3535.155 Million cell updates/sec

Title: US-10-614-282-1

Perfect score: 239

Sequence: 1 atcagtcacgacttgtaag.....tgtggtccggtgacaat 239

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_htc:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_htc:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: em\_gss\_hum:\*

18: em\_gss\_inv:\*

19: em\_gss\_pln:\*

20: em\_gss\_vrt:\*

21: em\_gss\_fun:\*

22: em\_gss\_mam:\*

23: em\_gss\_mus:\*

24: em\_gss\_pro:\*

25: em\_gss\_rod:\*

26: em\_gss\_phg:\*

27: em\_gss\_vrl:\*

28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	239	100.0	510	12	BI375946
2	37.2	15.6	1138	13	BQ144508
3	35.6	14.9	162	13	BM068571
4	35	14.6	496	10	BF024030

C	5	34.8	14.6	625	14	CD648340
C	6	34.8	14.6	1201	13	EX420094
C	7	34.6	14.5	563	28	AZ581975
C	8	34	14.2	670	29	LBAF025H06
C	9	34	14.2	670	29	LBAF025H09
C	10	33.8	14.1	507	29	EX137891
C	11	33.6	14.1	638	28	AZ983355
C	12	33.4	14.0	1026	11	AF227817
C	13	33.2	13.9	537	13	BQ743358
C	14	32.8	13.7	349	13	BU578178
C	15	32.8	13.7	438	10	AW133037
C	16	32.8	13.7	443	10	BE803253
C	17	32.8	13.7	446	12	BG725287
C	18	32.8	13.7	773	28	BH465834
C	19	32.8	13.6	955	13	BU901940
C	20	32.6	13.6	624	29	EX139003
C	21	32.6	13.6	852	14	CD649081
C	22	32.2	13.5	895	14	CD048532
C	23	32.2	13.5	565	13	BY432451
C	24	32.2	13.5	648	29	CE691717
C	25	32.2	13.5	664	10	BE941338
C	26	32.2	13.5	1201	9	AL573127
C	27	32	13.4	134	9	AU077131
C	28	32	13.4	608	12	BI561029
C	29	32	13.4	726	10	BG037177
C	30	32	13.4	775	12	BG616128
C	31	32	13.4	786	13	EX116922
C	32	31.8	13.3	731	28	BH992982
C	33	31.8	13.3	739	13	EX605323
C	34	31.8	13.3	776	29	CC532579
C	35	31.8	13.3	1267	28	CC324651
C	36	31.6	13.2	354	9	AU273213
C	37	31.6	13.2	653	28	AZ760934
C	38	31.6	13.2	739	14	CA413042
C	39	31.6	13.2	1077	13	BQ952602
C	40	31.4	13.1	364	9	AU082434
C	41	31.4	13.1	425	28	AZ404419
C	42	31.4	13.1	466	10	AW448888
C	43	31.4	13.1	549	10	BF657587
C	44	31.4	13.1	555	9	AU082410
C	45	31.4	13.1	654	14	CD233195

## ALIGNMENTS

### RESULT 1

BI375946

LOCUS

DEFINITION

RE63854.5prime RE Drosophila melanogaster

BI375946

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

BI375946 510 bp mRNA linear EST 01-AUG-2001  
Drosophila melanogaster cDNA clone RE63854 5, mRNA sequence.  
Drosophila melanogaster

GI:15071974

EST

Drosophila melanogaster (fruit fly)

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 510)

Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B.,

Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E.,

George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G.,

Misra, S., Mungall, C.J., Nunoo, J., Pacleb, J., Paragas, V., Park, S.,

Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celniker, S. and

Rubin, G.M.

BDGP/HMI RE Drosophila EST Project

Unpublished (2001)

Contact: Stapleton, M.

BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798



```

113 TCGGTAATATCTTTAGAAAGCAACAGCTAAGAACTATTTCAAGAACTGTGTGGCAAG 172
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
62  TTCCAAAATGATGAAAGAAATATCCCAACCCCTTGTNACGAGGTGTTTACAAG 121
QY      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
173 TGAAGGGTAGTTAGTGATACACCGGTTA 200
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
122 AGAGCAACTGTGGTTAAACACCGGTTA 149
QY      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 4
BF024030
LOCUS      496 bp mRNA linear EST 30-OCT-2000
DEFINITION PVP 300 L99-29 Litopenaeus vannamei cDNA similar to 16S rRNA, mRNA
sequence.
ACCESSION BF024030
VERSION BF024030.1 GI:11038174
KEYWORDS EST.
SOURCE Litopenaeus vannamei (Pacific white shrimp)
ORGANISM Litopenaeus vannamei
Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
Eumalacostraca; Eucarida; Decapoda; Dendrobranchiata; Penaeoidea;
Penaeidae; Litopenaeus.
REFERENCE 1 (bases 1 to 496)
AUTHORS Gross, P.S., Bartlett, T.C., Browdy, C.L., Chapman, R.W. and Warr, G.W.
TITLE Immune gene discovery by expressed sequence tag analysis of
hemocytes and hepatopancreas in the Pacific White Shrimp,
Litopenaeus vannamei, and the Atlantic White Shrimp, L. setiferus
Dev. Comp. Immunol. 25 (7), 565-577 (2001)
JOURNAL 11472779
MEDLINE
PUBMED
COMMENT Contact: Gross PS, Bartlett TC
Department of Biochemistry and Molecular Biology
Medical University of South Carolina
Box 250509, 173 Ashley Avenue, Charleston, SC 29425, USA
Tel: 846 792 9935
Fax: 843 792 4850
Email: gross@musc.edu.

FEATURES
source
LOCATION/Qualifiers
1..496
/organism="Litopenaeus vannamei"
/mol_type="mRNA"
/db_xref="taxon:6689"
/tissue_type="Hepatopancreas"
/clone_lib="L99-29"
/note="Vector: pTriplex2; Site_1: Sfi I; Site_2: Sfi I"

ORIGIN
Query Match 14.6%; Score 35; DB 10; Length 496;
Best Local Similarity 57.1%; Pred. No. 6.4;
Matches 56; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 115 CGGTAATATCTTTAGAAAGCAACAGCTAAGAACTATTTCAAGAACTGTGTGGCAAGTG 174
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
248 CGGTAATCTCTTTGAGAGTCCACATCGNAGGAAGGGTTGCGACCCGATGNTGAATN 307
QY      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
175 AAGGCTAGTATGATACACCGGTTATATCGAGTGGC 212
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
308 AAGGCTTCCTTATANTGCACCGTATATAAGGAGGGGC 345
QY      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 5
CD648340/c
LOCUS      625 bp mRNA linear EST 18-JUN-2003
DEFINITION AUF 102_H24_T7 Crassostrea virginica Gill Crassostrea virginica
cDNA 5' similar to Unknown, mRNA sequence.
ACCESSION CD648340
VERSION CD648340.1 GI:31904281
KEYWORDS EST.
SOURCE Crassostrea virginica (eastern oyster)
ORGANISM Crassostrea virginica
Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Ostreoida;
Ostreidae; Ostreidae; Crassostrea.

```

```

REFERENCE 1 (bases 1 to 625)
AUTHORS Peatman, E., Kucuktas, H., Li, P., He, C., Feng, J., Wei, X. and Liu, Z.
TITLE Differentially expressed oyster (Crassostrea virginica) genes after
exposure to mercury
JOURNAL Unpublished (2003)
COMMENT Contact: Liu ZJ
The Fish Molecular Genetics and Biotechnology Laboratory,
Department of Fisheries and Allied Aquacultures and Program of Cell
and Molecular Biosciences
Auburn University
203 Swingle Hall, Auburn University, Auburn, AL 36849, USA
Tel: 334 844 4054
Fax: 334 844 9208
Email: zliu@acesag.auburn.edu
Seq primer: M13 Reverse.
FEATURES
source
LOCATION/Qualifiers
1..625
/organism="Crassostrea virginica"
/mol_type="mRNA"
/db_xref="taxon:6565"
/clone_lib="Crassostrea virginica Gill"
/note="Organ: Gill; Vector: pSport1; Site_1: NotI; Site_2:
Sall"

ORIGIN
Query Match 14.6%; Score 34.8; DB 14; Length 625;
Best Local Similarity 57.3%; Pred. No. 7.9;
Matches 63; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 128 AGAAGCAACAGCTAAGAACTATTTCAAGAACTGTGTGGCAAGTGAAGGTAGTTAGT 187
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
345 AAAAAGCAATAGAAAAAACTATTTACACACACACACAAATGTAGCTTGTTACT 286
QY      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
188 GATACACCGGTTATATCGAGTGGCGGAAAGTGTGTCGGGTGGGACA 237
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
285 AAGAACTACTCTTTTGAAGTTCGATATCTGATGTTCTCGGTGACA 236
QY      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
BX420094/c
LOCUS      1201 bp mRNA linear EST 13-MAY-2003
DEFINITION BX420094 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
CSODF023YF10 5-PRIME, mRNA sequence.
ACCESSION BX420094
VERSION BX420094.1 GI:30638109
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1201)
AUTHORS Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 2290.f For
more information about this cluster, see
http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CSODF023DC05QPI&cluster=2290.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CSODF023DC05QPI.
FEATURES
source
LOCATION/Qualifiers
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODF023YF10"
/tissue_type="FETAL BRAIN"

```

(<http://www.jax.org/resources/documents/dnares/>). The DNA

	Query Match	14.2%	Score 34;	DB 29;	Length 670;
	Best Local Similarity	52.1%;	Pred. No. 14;		
	Matches	76;	Conservative	0;	Mismatches 70; Indels 0; Gaps 0;
QY	25	CAGGCAGCAGCGTGTGTGTGTATCGCCCAACGGAGTCGTGTTTTTCGTTTCGATACAGA	84		
Db	270	CATACACCGGACGTCGTGCTCCCTCGGCCAACATACGTGTGTGCGTGGCTTCCCGCGTA	211		



QY 85 TAAACCCACGTCGATAGCCCTCGACCGTCGGTAATATCTTAAAGCAACAGCTAA 144  
 Db 210 CCCAGCTCAGATCGAGTAGCTTCCAGGCCACCAGGGTCTTCGGTACCTCGCGAGGTCA 151  
 QY 145 AGAACTATTTCAAGAACTGTGTGGCA 170  
 Db 150 CCCAGTCTCCATGCACCTTTGTAGCA 125

RESULT 9  
 LBAF025H09/c  
 LOCUS LBAF025H09 670 bp DNA linear GSS 20-JUN-2003  
 DEFINITION Leishmania braziliensis GSS, clone LBAF025H09, genomic survey sequence.  
 ACCESSION BX542273.1 GI:32136112  
 VERSION GSS; genomic survey sequence.  
 KEYWORDS Leishmania braziliensis  
 SOURCE Leishmania braziliensis  
 ORGANISM Leishmania braziliensis  
 Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;  
 Leishmania; Leishmania braziliensis species complex.

REFERENCE 1  
 Laurentino, E.C., Ruiz, J.C. and Cruz, A.K.  
 GSS analysis of the Leishmania braziliensis genome  
 Unpublished  
 2 (bases 1 to 670)  
 Cruz, A.K.  
 Direct Submission  
 Submitted (17-JUN-2003) Cruz A.K., University of Sao Paulo,  
 Department of Molecular and Cell Biology, FMRP, Avenida  
 Bandeirantes 3900, Ribeirao Preto, SP, 14049-900, BRAZIL  
 Clone requests: akruz@fmrp.usp.br.  
 COMMENT Location/Qualifiers  
 FEATURES  
 source  
 1..670  
 /organism="Leishmania braziliensis"  
 /mol\_type="genomic DNA"  
 /strain="MHOM/BR/75/M2904"  
 /db\_xref="taxon:5660"  
 /clone="LBAF025H09"

ORIGIN  
 Query Match 14.2%; Score 34; DB 29; Length 670;  
 Best Local Similarity 52.1%; Pred. No. 14;  
 Matches 76; Conservative 0; Mismatches 70; Indels 0; Gaps 0;  
 QY 25 CAGGACGACGTCGTCGTCATCGCAACGGAGTCGTTTTCGGTTCGATACAGA 84  
 Db 270 CATACCGCGACGTCGTCGCCCTCGGCCAACATACGTCGTCGTCGCGGTA 211  
 QY 85 TAAACCCACGTCGATAGCCCTCGACCGTCGGTAATATCTTAAAGCAACAGCTAA 144  
 Db 210 CCCAGCTCAGATCGAGTAGCTTCCAGGCCACCAGGGTCTTCGGTACCTCGCGAGGTCA 151  
 QY 145 AGAACTATTTCAAGAACTGTGTGGCA 170  
 Db 150 CCCAGTCTCCATGCACCTTTGTAGCA 125

RESULT 10  
 BX137891  
 LOCUS BX137891 507 bp DNA linear GSS 13-MAR-2003  
 DEFINITION Danio rerio genomic clone DKEY-86F3, genomic survey sequence.  
 ACCESSION BX137891  
 VERSION BX137891.1 GI:27969207  
 KEYWORDS GSS.  
 SOURCE Danio rerio (zebrafish)  
 ORGANISM Danio rerio  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
 Cypriniformes; Cyprinidae; Danio.  
 REFERENCE 1 (bases 1 to 507)  
 HUMPHRAY, S.J., HUCKLE, E. and DURHAM, J.L.

# TITLE JOURNAL

## COMMENT

# FEATURES source

1..507  
 /organism="Danio rerio"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7955"  
 /clone="DKEY-86F3"  
 /tissue\_type="testis"  
 /note="vector plndigoBAC-536"

## ORIGIN

Query Match 14.1%; Score 33.8; DB 29; Length 507;  
 Best Local Similarity 51.0%; Pred. No. 15;  
 Matches 80; Conservative 0; Mismatches 77; Indels 0; Gaps 0;  
 QY 53 AACGGAGTCGTGTTTTCGGTTCGATACAGATAAAACCCACGTCGATAGCCCTCGACCG 112  
 Db 336 AATCGGTGCACTATTTCACCGTTGTCATTTAAGGACAGATACATAGACTCTTTTG 395  
 QY 113 TCGGTATATTTCTAGAAAGCAACAGCTTAAAGAACTATTTCAAGAACTGTGTGGCAAG 172  
 Db 396 TGTTTATTTTATTAATAAAATAACAATGTTAGATAAAATATACAAATATTTGGGAAA 455  
 QY 173 TGAAGGTAGTAGTAGTACACCGGTTATATCGCAGT 209  
 Db 456 AGTCAGTCAATTATAGCTATATGGGTTTATTGGAAT 492

# RESULT 11

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## FEATURES

## source

## 1..638

## /organism="Mus musculus"

## /mol\_type="genomic DNA"

## /strain="C57BL/6J"

## Location/Qualifiers

## High quality sequence stop: 638.

## Location/Qualifiers

## 1..638

## /organism="Mus musculus"

## /mol\_type="genomic DNA"

## /strain="C57BL/6J"

## Location/Qualifiers

## High quality sequence stop: 638.

## Location/Qualifiers

## 1..638

## /organism="Mus musculus"

## /mol\_type="genomic DNA"

## /strain="C57BL/6J"

# Direct Submission

Submitted (13-MAR-2003) The Sanger Institute, Wellcome Trust Genome  
 Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humqurys@sanger.ac.uk Unpublished  
 This sequence was generated from the T7 end of BAC 86F3. 86F3 is  
 part of the Daniokey BAC Library created by R. Plasterk and N.V.  
 Keygene. Further details:  
 http://www.sanger.ac.uk/Projects/D\_rerio/.

# Location/Qualifiers

## 1..507

## /organism="Danio rerio"

## /mol\_type="genomic DNA"

## /db\_xref="taxon:7955"

## /clone="DKEY-86F3"

## /tissue\_type="testis"

## /note="vector plndigoBAC-536"

## Location/Qualifiers

## 1..507

## /organism="Danio rerio"

## /mol\_type="genomic DNA"

## /db\_xref="taxon:7955"

## /clone="DKEY-86F3"

## /tissue\_type="testis"

## /note="vector plndigoBAC-536"

## Location/Qualifiers

## 1..507

## /organism="Danio rerio"

## /mol\_type="genomic DNA"

## /db\_xref="taxon:7955"

## /clone="DKEY-86F3"

## /tissue\_type="testis"

## /note="vector plndigoBAC-536"

## Location/Qualifiers

## 1..507

## /organism="Danio rerio"

## /mol\_type="genomic DNA"

## /db\_xref="taxon:7955"

## /clone="DKEY-86F3"

## /tissue\_type="testis"

## /note="vector plndigoBAC-536"

## Location/Qualifiers

## 1..507

## /organism="Danio rerio"

## /mol\_type="genomic DNA"

## /db\_xref="taxon:7955"

## /clone="DKEY-86F3"

## /tissue\_type="testis"

## /note="vector plndigoBAC-536"

## Location/Qualifiers

## 1..507

## /organism="Danio rerio"

## /mol\_type="genomic DNA"

## /db\_xref="taxon:7955"

## /clone="DKEY-86F3"

## /tissue\_type="testis"

## /note="vector plndigoBAC-536"

## Location/Qualifiers

## 1..507

## /organism="Danio rerio"

## /mol\_type="genomic DNA"

## /db\_xref="taxon:7955"

## /clone="DKEY-86F3"

# Location/Qualifiers

## 1..507

## /organism="Danio rerio"

## /mol\_type="genomic DNA"

## /db\_xref="taxon:7955"

## /clone="DKEY-86F3"

## /tissue\_type="testis"

## /note="vector plndigoBAC-536"

## Location/Qualifiers

## 1..507

## /organism="Danio rerio"

## /mol\_type="genomic DNA"

## /db\_xref="taxon:7955"

## /clone="DKEY-86F3"

## /tissue\_type="testis"

## /note="vector plndigoBAC-536"

## Location/Qualifiers

## 1..507

## /organism="Danio rerio"

## /mol\_type="genomic DNA"

## /db\_xref="taxon:7955"

## /clone="DKEY-86F3"

## /tissue\_type="testis"

## /note="vector plndigoBAC-536"

## Location/Qualifiers

## 1..507

## /organism="Danio rerio"

## /mol\_type="genomic DNA"

## /db\_xref="taxon:7955"

## /clone="DKEY-86F3"

## /tissue\_type="testis"

## /note="vector plndigoBAC-536"

## Location/Qualifiers

## 1..507

## /organism="Danio rerio"

## /mol\_type="genomic DNA"

## /db\_xref="taxon:7955"

## /clone="DKEY-86F3"

## /tissue\_type="testis"

## /note="vector plndigoBAC-536"

## Location/Qualifiers

## 1..507

## /organism="Danio rerio"

## /mol\_type="genomic DNA"

## /db\_xref="taxon:7955"

## /clone="DKEY-86F3"

## /tissue\_type="testis"

## /note="vector plndigoBAC-536"

## Location/Qualifiers

## 1..507

## /organism="Danio rerio"

## /mol\_type="genomic DNA"

## /db\_xref="taxon:7955"

## /clone="DKEY-86F3"

## /tissue\_type="testis"

## /note="vector plndigoBAC-536"

## Location/Qualifiers

## 1..507

## /organism="Danio rerio"

## /mol\_type="genomic DNA"

# Location/Qualifiers

## 1..507

## /organism="Danio rerio"

## /mol\_type="genomic DNA"

## /db\_xref="taxon:7955"

## /clone="DKEY-86F3"

## /tissue\_type="testis"

## /note="vector plndigoBAC-536"

## Location/Qualifiers

## 1..507

## /organism="Danio rerio"

## /mol\_type="genomic DNA"

## /db\_xref="taxon:7955"

## /clone="DKEY-86F3"

## /tissue\_type="testis"

```

/db_xref="taxon:10090"
/clone="UUGC2M0264C01"
/sex="female"
/lab_host="E. coli strain XL10-Gold, Tl-resistant, F-"
/clone_lib="Mouse 10kb plasmid UUGC2M library"
/notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN
Query Match 14.1%; Score 33.6; DB 28; Length 638;
Best Local Similarity 61.4%; Pred. No. 19;
Matches 54; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 130 AAAGCAACAGCTAAGAACTATTTCAGAACTGTGGCAAGTGAAGGTTAGTGA 189
DB 310 AAAGCACTTCAAAATCTATTAACTATGTGCTAATTAAGAAATTTAGATA 369

QY 190 TACACGGTTATATCGGAGTGGCGAGAA 217
DB 370 TATTCGTGTTATTTGGAAACCTATA 397

RESULT 12
AF227817/c
LOCUS
DEFINITION
Ictalurus punctatus mitochondrial long-chain enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase mRNA, partial sequence.
ACCESSION
AF227817.1 GI:27883582
VERSION
SOURCE
Ictalurus punctatus (channel catfish)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Siluriformes; Ictaluridae; Ictalurus.
REFERENCE
1 (bases 1 to 1026)
AUTHORS
Kim,S., Li,P., Zheng,X., Dunham,R.A. and Liu,Z.
TITLE
Gene expression in the muscles of young and mature channel catfish (Ictalurus punctatus) as analyzed by expressed sequence tags and gene filters.
JOURNAL
Fish Physiol. Biochem. (2003) In press
REFERENCE
2 (bases 1 to 1026)
AUTHORS
Liu,Z. and Kim,S.
TITLE
Direct Submission
JOURNAL
Submitted (21-JAN-2000) Fisheries and Allied Aquacultures and the Program of Cell and Molecular Biosciences, Auburn University, 203 Swingle Hall, Auburn, AL 36849, USA
FEATURES
Location/Qualifiers
1..1026
/organism="Ictalurus punctatus"
/mol_type="mRNA"
/db_xref="taxon:7998"
/tissue_type="muscle"
misc_feature
<1..>1026
/notes="similar to mitochondrial long-chain enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase; coding region"

```

```

ORIGIN
Query Match 14.0%; Score 33.4; DB 11; Length 1026;
Best Local Similarity 65.3%; Pred. No. 25;
Matches 49; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 110 CGTCGCGTAATATCTTAGAAAGCAACAGCTAAGAACTATTTCAGAACTGTGGC 169
DB 911 CCCTTCTCTAGAACTCTTAGATCACAGACCTTCGGAACGATTTTAGAACTGTGAGT 852

QY 170 AAGTGAAGGTTAGTT 184
DB 851 AAGAGAAATGTCGGT 837

RESULT 13
BQ743358
LOCUS
DEFINITION
WHE4103_A10_B19S Wheat salt-stressed root cDNA library Triticum aestivum cDNA clone WHE4103_A10_B19, mRNA sequence.
ACCESSION
BQ743358
VERSION
BQ743358.1 GI:21890145
KEYWORDS
EST.
SOURCE
Triticum aestivum (bread wheat)
ORGANISM
Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae; Triticum.
REFERENCE
1 (bases 1 to 537)
AUTHORS
Anderson,O.D., Akhunov,E., Chao,S., Crossman,C., Deal,K., Dvorak,J., Lazo,G.R., Pham,J., Rausch,C.J., Wilson,C. and Woo,J.
TITLE
The structure and function of the expressed portion of the wheat genomes - Salt-stressed root cDNA library
JOURNAL
Unpublished (2002)
COMMENT
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oanderson@pw.usda.gov
Sequences have been trimmed to remove vector sequence and low quality sequence with phred score less than 20
Seq primer: SK primer.
FEATURES
Location/Qualifiers
1..537
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHE4103_A10_B19"
/tissue_type="Roots"
/dev_stage="Full tillering"
/lab_host="E. coli SOLR"
/clone_lib="Wheat salt-stressed root cDNA library"
/notes="Vector: Lambda Uni-ZAP XR, excised phagemid pBluescript SK(-); Site 1: EcoRI; Site 2: XhoI; Hydroponic plants grown to full tillering stage were treated with 150 mM NaCl for either 12 hours or 7 days. Root tissues of the plants subjected to both types of treatment were collected separately at University of California, Davis (E. Akhunov and K. Deal in J. Dvorak's Lab). Total RNA was prepared separately from the two samples (12h and 7day treatments), and equal amount of RNA was then pooled. PolyA RNA was purified from the pooled RNA, a cDNA library was made, and SK(-) phagemids in J. Dvorak's lab (E. Akhunov, J. Dvorak) at the University of California, Davis. Colony plating, plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."
ORIGIN

```

Query Match 13.9%; Score 33.2; DB 13; Length 537;  
 Best Local Similarity 59.6%; Pred. No. 24;  
 Matches 56; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 23 CGCAGGACGACGTCGTCTCATCCCAACGGAGTCGTGTTTCGGTTCGATACA 82  
 |||  
 Db 74 CGCGGAAGCAGTCGTCTGCCATCCGCCGCGTGTCTCGCGTCCGAACAGAAAGC 133  
 |||

QY 83 GATAAACCACGTCGATAGCCCTCGACCGTCGC 116  
 |||  
 Db 134 ACGGTACAGCAGGAGGAGGAGCCCGGTCGTGC 167  
 |||

RESULT 14  
 BU578178  
 LOCUS  
 DEFINITION sar48g12.y1 Gm-cl074 Glycine max cDNA clone SOYBEAN CLONE ID:  
 Gm-cl074-5711 5' similar to TR:080921 080921 PUTATIVE RNA-BINDING  
 PROTEIN. ;, mRNA sequence.

ACCESSION BU578178  
 VERSION BU578178.1 GI:23057504  
 SOURCE Glycine max (soybean)  
 ORGANISM Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
 Glycine.

1 (bases 1 to 349)  
 REFERENCE Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V.,  
 Khanna,A., Bolla,B., Marra,M., Hillier,B., Kucaba,T., Martin,J.,  
 Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,  
 Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,  
 Schurr,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,  
 McCann,R., Waterston,R. and Wilson,R.  
 Public Soybean EST Project  
 Unpublished (1999)

TITLE Public Soybean EST Project  
 JOURNAL  
 COMMENT Contact: Shoemaker R/Public Soybean EST Project  
 Public Soybean EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu

This clone is available through: ResGen, Invitrogen Corp. 2130  
 South Memorial Parkway Huntsville, AL 35801 For further information  
 call: (800)-533-4363 or contact: ccu@resgen.com web site:  
 www.resgen.com

Seq primer: -40RP from Gibco  
 High quality sequence stop: 295.

FEATURES  
 source

1. .349  
 /organism="Glycine max"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:3847"  
 /clone="SOYBEAN CLONE ID: Gm-cl074-5711"  
 /tissue\_type="seedlings induced for HR (hypersensitive  
 response)"  
 /dev\_stage="9-11 day old"  
 /lab\_host="DH10B"  
 /clone\_lib="Gm-cl074"  
 /note="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2:  
 XhoI; The cDNA library was constructed from mRNA isolated  
 from 9-11 day old seedlings that were induced for HR  
 (hypersensitive response) by vacuum infiltrating plant  
 tissue with Pseudomonas syringae pv. glycinea carrying the  
 avrB gene (Genetics 141:1597-1604). plant tissue (expanded  
 unifoliate leaves) was collected at 2, 4, 8, 12, 24, 36,  
 and 53 hrs after inoculation and their mRNA pooled equally  
 for cDNA construction. The library was prepared using the  
 Stratagene pBluescript II SK(+) library construction kit.  
 Complementary DNA was synthesized from mRNA using a primer  
 consisting of a poly(dT) sequence with an XhoI restriction

site. EcoRI adaptors were ligated to the blunt-ended cDNA  
 fragments followed by XhoI digestion. The cDNA insert is  
 protected from XhoI digestion via methylation during first  
 strand synthesis. The cDNA fragments were directionally  
 cloned into the EcoRI-XhoI restriction site of the  
 pBluescript vector. The ligated cDNA fragments were  
 transformed into E.coli Electromax DH10B host cells. Plant  
 care, inoculations, and library construction were  
 performed by Steve Clough (Lila Vodkin lab, University of  
 Illinois)."

ORIGIN

Query Match 13.7%; Score 32.8; DB 13; Length 349;  
 Best Local Similarity 61.9%; Pred. No. 28;  
 Matches 52; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 124 TCTTAGAAGCAACAGCTAAAGAACTATTTCAGAACTGTGTGGCAAGTCAAGGTAGT 183  
 |||  
 Db 105 TCTTAGCAACGTACTACTACAGAAAGCTTCGAGAGAGTTTCGAAAGTTGTGTAAGT 164  
 |||

QY 184 TAGTGATACACCGCTTATATCGGA 207  
 |||  
 Db 165 TGTTCATGCAAGGCTTGTAACTGA 188  
 |||

RESULT 15  
 AW133037  
 LOCUS

DEFINITION AW133037 438 bp mRNA linear EST 30-NOV-2001  
 sel3e10.y1 Gm-cl013 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:  
 Gm-cl013-3427 5' similar to TR:Q9ZUR6 Q9ZUR6 PUTATIVE RNA BINDING  
 PROTEIN. ;, mRNA sequence.

ACCESSION AW133037  
 VERSION AW133037.1 GI:6134644  
 SOURCE EST  
 ORGANISM Glycine max (soybean)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
 Glycine.

1 (bases 1 to 438)

REFERENCE Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V.,  
 Khanna,A., Bolla,B., Marra,M., Hillier,B., Kucaba,T., Martin,J.,  
 Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,  
 Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,  
 Schurr,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,  
 McCann,R., Waterston,R. and Wilson,R.  
 Public Soybean EST Project  
 Unpublished (1999)

TITLE Public Soybean EST Project  
 JOURNAL  
 COMMENT Contact: Shoemaker R/Public Soybean EST Project  
 Public Soybean EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu

This clone is available through: ResGen, Invitrogen Corp. 2130  
 South Memorial Parkway Huntsville, AL 35801 For further information  
 call: (800)-533-4363 or contact via email: ccu@resgen.com  
 Insert Length: 772 Std Error: 0.00  
 Seq primer: -40RP from Gibco

High quality sequence stop: 271.  
 Location/Qualifiers

FEATURES  
 source

1. .438  
 /organism="Glycine max"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:3847"  
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl013-3427"  
 /tissue\_type="Whole seedlings, 2-3 week old seedlings,  
 greenhouse grown"  
 /lab\_host="Xl10-Gold"  
 /clone\_lib="Gm-cl013"  
 /note="Vector: pBluescript II XR; Site 1: EcoRI; Site 2:

XhoI. This cDNA library was constructed from mRNA isolated from whole seedlings of 2-3 week old greenhouse grown plants. The cDNA library was prepared using the Stratagene pBluescript II XR cDNA library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into XL10-Gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erpeliding."

ORIGIN

Query Match 13.7%; Score 32.8; DB 10; Length 438;  
 Best Local Similarity 61.9%; Pred. No. 30;  
 Matches 52; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY	124	TCCTAGAAAGCAACAGCTAAGAACTATTTCAGAACTGTGTGGCAAGTGAAGGTAGT	183
Db	213		
QY	184	TAGTCATACACCGGTTATATCGGA	207
Db	273		
		TCCTCATGCAAGGGTTGTAAGTGA	296

Search completed: September 18, 2004, 13:16:48  
 Job time : 2023.88 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 2004, 09:46:06 ; Search time 1166.45 Seconds  
(without alignments)  
9066.559 Million cell updates/sec

Title: US-10-614-282-2  
Perfect score: 244  
Sequence: 1 agccgcggggttaaatgc.....accaggtcaactcagtgacag 244

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_ov.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_ov.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	244	100.0	102717	9 AC004079	AC004079 Homo sapi
C 2	242.4	99.3	170237	9 AC142285	AC142285 Pan trogl
C 3	240.8	98.7	191971	9 AC146102	AC146102 Pan trogl
C 4	204.8	83.9	128851	2 AC129963	AC129963 Sus scrofa
C 5	193.2	81.6	112570	10 AC116662	AC116662 Rattus no
C 6	193.2	81.6	234603	2 AC097087	AC097087 Rattus no
C 7	199.2	81.6	298853	2 AC122669	AC122669 Rattus no
C 8	184	75.4	200633	2 AC015583	AC015583 Mus muscu
C 9	184	75.4	236214	2 AC091106	AC091106 Mus muscu
C 10	184	75.4	236829	2 AC123672	AC123672 Mus muscu
C 11	172.6	70.7	6118	6 AX344703	AX344703 Sequence
C 12	135.2	55.4	6118	6 AX344702	AX344702 Sequence
C 13	62	25.4	2529	9 BC032547	BC032547 Homo sapi
C 14	53.8	22.0	2216	10 MUSHOX161	M22115 Mouse ERA-1
C 15	51	20.9	443	10 MUSHOX161	M20214 Mus musculu
C 16	44	18.0	209926	2 AC145789	AC145789 Silurana
C 17	41.4	17.0	160516	10 AL928943	AL928943 Mouse DNA
C 18	40.8	16.7	204697	2 AC120995	AC120995 Rattus no
C 19	40.8	16.7	236070	2 AC112571	AC112571 Rattus no
C 20	38.8	15.9	158174	8 AP003414	AP003414 Oryza sat
C 21	38.8	15.9	159761	8 AP003921	AP003921 Oryza sat
C 22	38	15.6	41722	9 AL359316	AL359316 Human DNA
C 23	38	15.6	302250	6 AX335531	AX335531 Sequence
C 24	38	15.6	324604	9 U82671	U82671 Homo sapien
C 25	37.8	15.5	77457	1 AF210249	AF210249 Streptomy
C 26	37.6	15.4	244549	2 AC147476	AC147476 Mus muscu
C 27	37.4	15.3	215259	2 AC118349	AC118349 Rattus no
C 28	37.2	15.2	719	11 PM7B12B	PM7B12B Penicilli
C 29	37.2	15.2	3050	6 AX746737	AX746737 Sequence
C 30	37.2	15.2	3050	9 AK091024	AK091024 Homo sapi
C 31	37.2	15.2	174367	10 AL512597	AL512597 Mouse DNA
C 32	37	15.2	16115	1 AE007075	AE007075 Mycobacte
C 33	37	15.2	63155	6 AX646257	AX646257 Sequence
C 34	37	15.2	63155	9 AB065664	AB065664 Homo sapi
C 35	37	15.2	172953	9 AC067721	AC067721 Homo sapi
C 36	37	15.2	306050	1 BX248341	BX248341 Mycobacte
C 37	37	15.2	348247	15 BX842579	BX842579 Mycobacte
C 38	36.8	15.1	1023	10 MMH0X16R	X06024 Mouse Hox-1
C 39	36.6	15.0	172945	9 AC007220	AC007220 Homo sapi
C 40	36.6	15.0	185908	9 AC136297	AC136297 Homo sapi
C 41	36.6	15.0	206894	10 AC109152	AC109152 Mus muscu
C 42	36.6	15.0	253850	2 AC096825	AC096825 Rattus no
C 43	36.6	15.0	257407	2 AC109870	AC109870 Rattus no
C 44	36.6	15.0	258319	2 AC127770	AC127770 Rattus no
C 45	36.6	15.0	266342	2 AC109963	AC109963 Rattus no

ALIGNMENTS

RESULT 1  
AC004079/c  
LOCUS AC004079 102717 bp DNA linear PRI 26-APR-2003  
DEFINITION Homo sapiens PAC clone RPI-167F23 from 7, complete sequence.  
ACCESSION AC004079  
VERSION AC004079.1 GI:2822174  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 102717)  
AUTHORS Sulston, J.E. and Waterston, R.  
TITLE Toward a complete human genome sequence  
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)



```

repeat_region      /rpt_family="Alu"
variation           8544_..8579
                   /rpt_family="AT-rich"
                   8760
                   /alleles="A"
                   /alleles="G"
                   /db_xref="dbSNP:2465263"
repeat_region      8991_..9300
variation           /rpt_family="Alu"
                   9745
                   /alleles="G"
                   /alleles="C"
                   /db_xref="dbSNP:2522829"
repeat_region      10140_..10406
variation           /rpt_family="L1"
                   10407_..10702
                   /rpt_family="Alu"
                   10602
                   /alleles="G"
                   /alleles="A"
                   /db_xref="dbSNP:2462902"
variation           10656
                   /alleles="G"
                   /alleles="A"
                   /db_xref="dbSNP:2254463"
repeat_region      10703_..10790
variation           /rpt_family="L1"
                   11252_..11503
                   /rpt_family="MIR"
                   11267
                   /alleles="C"
                   /alleles="G"
                   /db_xref="dbSNP:2522822"
repeat_region      11814_..11854
variation           /rpt_family="MIR"
                   12037_..12443
                   /rpt_family="Alu"
                   12465
                   /alleles="C"
                   /alleles="T"
                   /db_xref="dbSNP:2428416"

Query Match      100.0%; Score 244; DB 9; Length 102717;
Best Local Similarity 100.0%; Pred. No. 2,2e-60;
Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCGCGGGGATTAAATGCCACTAAACGGTGATCCATCATCTCGGAAGCGGCACAACT 60
Db 78551 AGCGCGGGGATTAAATGCCACTAAACGGTGATCCATCATCTCGGAAGCGGCACAACT 78492
QY 61 TTTCAGAGGCTCAGCCATTGGCTGACACCGTCACGTCGCCCTCTCCAGCGTCCTCCG 120
Db 78491 TTTCAGAGGCTCAGCCATTGGCTGACACCGTCACGTCGCCCTCTCCAGCGTCCTCCG 78432
QY 121 CCCTCCGCCGCCCTCTTGGCGACTGACATTCATATCATTTTCTTCGCGGCCCAT 180
Db 78431 CCCTCCGCCGCCCTCTTGGCGACTGACATTCATATCATTTTCTTCGCGGCCCAT 78372
QY 181 GGAGAGTGAAGAAGTTGGACAGTACCGCGGCTTCGAGGACCAAGGTCACTAGTG 240
Db 78371 GGAGAGTGAAGAAGTTGGACAGTACCGCGGCTTCGAGGACCAAGGTCACTAGTG 78312
QY 241 ACAG 244
Db 78311 ACAG 78308

```

```

RESULT 2
AC142285/C AC142285 170237 bp DNA linear PRI 26-SEP-2003
LOCUS Pan troglodytes BAC clone RP43-169B4 from 7, complete sequence.
DEFINITION AC142285
ACCESSION AC142285
VERSION AC142285.1 GI:29294404

```

```

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

```

HTG.
Pan troglodytes (chimpanzee)
Pan troglodytes
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
1 (bases 1 to 170237)
Belter,E. and Maupin,R.
The sequence of Pan troglodytes BAC clone RP43-169B4
Unpublished (2001)
2 (bases 1 to 170237)
Sulston,J.E. and Wilson,R.
Sequencing of Pan troglodytes
Unpublished (2001)
3 (bases 1 to 170237)
Wilson,R.K.
Direct Submission
Submitted (27-MAR-2003) Genetics, Genome Sequencing Center, 4444
Forest Park Parkway, St. Louis, MO 63108, USA
4 (bases 1 to 170237)
Wilson,R.
Direct Submission
Submitted (26-SEP-2003) Department of Genetics, Washington
University, 444 Forest Park Avenue, St. Louis, Missouri 63108, USA
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
----- Summary Statistics
-----
Center project name: C_PT169B04
-----

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu

SOURCE INFORMATION:  
The RPCI-43 BAC Library has been constructed by Chung-Li Shu. DNA was isolated from white blood cells obtained from a male chimpanzee (Pan troglodytes, 'Clint', Yerkes #C0471; birthdate: 6-6-80). The clone and detailed information can be obtained from Resgen (http://www.resgen.com) or Pieter de Jong and co-workers at http://www.bacpac.chori.org.

NEIGHBORING SEQUENCE INFORMATION:  
This sequence is the entire insert of the clone.

Discrepant bases between whole genome shotgun data and clone sequence.

Data from RP43-21L13 was used to finish this clone.

```

FEATURES
source
Location/Qualifiers
1..170237
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/chromosome="7"
/map="7"

```

```

/clone="RP43-159B4"
/clone_lib="RPCI-43"

ORIGIN
Query Match          99.3%; Score 242.4; DB 9; Length 170237;
Best Local Similarity 99.6%; Pred. No. 6.3e-60;
Matches 243; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACGCCGGGGATTAAATGCCACTAAACCGGTGATCCATCACTGGCGGAAGCCGGCAAACT 60
DB 119853 AGCGCGGGGATTAAATGCCACTAAACCGGTGATCCATCACTGGCGGAAGCCGGCAAACT 119794

QY 61 TTTCGAGGAGGCTCAGCCATGGCTGACACCGCTCAGTCGCCCTCTCCAGGTCCTCCG 120
DB 119793 TTTCGAGGAGGCTCAGCCATGGCTGACACCGCTCAGTCGCCCTCTCCAGGTCCTCCG 119734

QY 121 CCTCCCGCGCCCGCCCTCTTGGCGACATGTACATTCATATCATTTTCTTCCGGGCCCAAT 180
DB 119733 CCTCCCGCGCCCGCCCTCTTGGCGACATGTACATTCATATCATTTTCTTCCGGGCCCAAT 119674

QY 181 GGAGGAAGTGAGAAAGTTGGACAGTCAGTCGCCGGGTTGCGAGGACCGGTCATCAGTG 240
DB 119673 GGAGGAAGTGAGAAAGTTGGACAGTCAGTCGCCGGGTTGCGAGGACCGGTCATCAGTG 119614

QY 241 ACAG 244
DB 119613 ACAG 119610

RESULT 3
AC146102/c
LOCUS AC146102 191971 bp DNA linear PRI 18-DEC-2003
DEFINITION Pan troglodytes BAC clone RP43-21L13 from 7, complete sequence.
ACCESSION AC146102
VERSION AC146102.2 GI:38230057
KEYWORDS HTG.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
1 (bases 1 to 191971)
The sequence of Pan troglodytes BAC clone RP43-21L13
Unpublished (2001)
Wilson, R.
2 (bases 1 to 191971)
Sequencing of Pan troglodytes
Unpublished (2001)
Wilson, R.K.
3 (bases 1 to 191971)
Wilson, R.K.
Direct Submission
Submitted (01-AUG-2003) Genetics, Genome Sequencing Center, 4444
Forest Park Parkway, St. Louis, MO 63108, USA
4 (bases 1 to 191971)
Wilson, R.K.
Direct Submission
Submitted (09-NOV-2003) Genetics, Genome Sequencing Center, 4444
Forest Park Parkway, St. Louis, MO 63108, USA
5 (bases 1 to 191971)
Wilson, R.
Direct Submission
Submitted (18-DEC-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Nov 9, 2003 this sequence version replaced gi:33387055.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
----- Summary Statistics
Center project name: C_PT021L13
-----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
http://genome.wustl.edu

SOURCE INFORMATION:
The RPCI-43 BAC Library has been constructed by Chung-Li Shu. DNA
was isolated from white blood cells obtained from a male chimpanzee
(Pan troglodytes, 'Clint', Yerkes #C0471; birthdate: 6-6-80). The
clone and detailed information can be obtained from ResGen
(http://www.resgen.com) or Pieter de Jong and co-workers at
http://www.bacpac.chori.org.

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone.
Location/Qualifiers
1. 191971
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/chromosome="7"
/map="7"
/clone="RP43-21L13"
/clone_lib="RPCI-43"
22205..82239
/note="Sequence derived from PCR product of project DNA."
156808..156822
/note="Sequence derived from one plasmid subclone."
157072..157109
/note="Sequence derived from one plasmid subclone."
159887..160183
/note="Sequence derived from one plasmid subclone."
177866..177987
/note="Sequence derived from one plasmid subclone."
178716..178743
/note="Sequence derived from one plasmid subclone."

FEATURES
Source
misc_feature
22205..82239
/note="Sequence derived from PCR product of project DNA."
unsure
156808..156822
/note="Sequence derived from one plasmid subclone."
unsure
157072..157109
/note="Sequence derived from one plasmid subclone."
unsure
159887..160183
/note="Sequence derived from one plasmid subclone."
unsure
177866..177987
/note="Sequence derived from one plasmid subclone."
unsure
178716..178743
/note="Sequence derived from one plasmid subclone."

ORIGIN
Query Match          98.7%; Score 240.8; DB 9; Length 191971;
Best Local Similarity 99.2%; Pred. No. 1.9e-59;
Matches 242; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ACGCCGGGGATTAAATGCCACTAAACCGGTGATCCATCACTGGGAAGCCGGCAAACT 60
DB 99935 AGCGCGGGGATTAAATGCCACTAAACCGGTGATCCATCACTGGGAAGCCGGCAAACT 99876

QY 61 TTTCGAGGAGGCTCAGCCATGGCTGACACCGCTCAGTCGCCCTCTCCAGGTCCTCCG 120
DB 99875 TTTCGAGGAGGCTCAGCCATGGCTGACACCGCTCAGTCGCCCTCTCCAGGTCCTCCG 99816

QY 121 CCTCCCGCGCCCGCCCTCTTGGCGACATGTACATTCATATCATTTTCTTCCGGGCCCAAT 180
DB 99815 CCTCCCGCGCCCGCCCTCTTGGCGACATGTACATTCATATCATTTTCTTCCGGGCCCAAT 99756

QY 181 GGAGGAAGTGAGAAAGTTGGACAGTCAGTCGCCGGGTTGCGAGGACCGGTCATCAGTG 240
DB 99755 GGAGGAAGTGAGAAAGTTGGACAGTCAGTCGCCGGGTTGCGAGGACCGGTCATCAGTG 99696

QY 241 ACAG 244
```



```

Db      99695 ACAG 99692

|||||
----- Project Information
Center project name: L24464
Center clone name: 485_M_13

* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 6913: contig of 6913 bp in length
* 6914 7013: gap of 100 bp
* 7014 10891: contig of 3878 bp in length
* 10892 10991: gap of 100 bp
* 10992 71101: contig of 60110 bp in length
* 71102 71201: gap of 100 bp
* 71202 128851: contig of 57650 bp in length.

FEATURES             source
1..128851
/organism="Sus scrofa"
/mol_type="genomic DNA"
/db_xref="taxon:9823"
/clone="RP44-485M13"
/clone_lib="RPC1-44 Male Porcine BAC"

ORIGIN
Query Match      83.9%   Score 204.8; DB 2; Length 128851;
Best Local Similarity 92.2%; Pred. No. 6.1e-49;
Matches 226; Conservative 0; Mismatches 18; Indels 1; Gaps 1;

Qy 1 AGCCGCGGGATTAAATGCCACTTAAACCGGTGATCCATCATCTCGGAAGCGGCAAACT 60
Db 115286 AGCCGCGGGATTAAATGCCACTGAAAGGTGATCATCACCGCAGGAGCGCAAACT 115345

Qy 61 TTTCAGAGAGCTCAGCATTGGCTGTGAC-ACCGTCAGTGCCTCTCTCCAGGCTCTCTCC 119
Db 115346 TTTCAGAGAGCTCAGCATTGGCTGTGAGAGTCACTGCTCCCTCTCCAGGCTCTCTCC 115405

Qy 120 GCCCTCCCGCCCCCTCTTGGCGACTGTACATTCATCATATTTCTTCGCGGCCCA 179
Db 115406 GCCCTCCCGCCCCCTCTTGGCGACTGTACATTCATCATATTTCTTCGCGGCCCA 115465

Qy 180 TGGAGGAAGTGAGAAAGTTGGCAGTCAGCCGGGTTTCGAGGACCGAGTCACTCAGT 239
Db 115466 TGGAGGAAGTGAGAAAGTTGGCAGTCAGTCAGGCTTGCAGGACCCGCTCACTCAGT 115525

Qy 240 GACAG 244
Db 115526 GACAG 115530

RESULT 5
AC116662
LOCUS      AC116662      112570 bp      DNA      linear      ROD 02-MAY-2002
DEFINITION Rattus norvegicus, clone RP31-256A2, complete sequence.
ACCESSION  AC116662
VERSION     AC116662.1 GI:19852147
KEYWORDS    HTG.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
            Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 112570)
AUTHORS    Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE      Rattus norvegicus, clone RP31-256A2
JOURNAL    Unpublished
REFERENCE   2 (bases 1 to 112570)
AUTHORS    Dewart,K., Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A.,
            Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T.,
            Boguslavskiy,L., Boukhgalter,B., Brown,A., Camarata,J.,
            Gargyala,S., Gordon,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliiev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kellis,C., Lander,E., Allen,A., Linton,L.,
            Mammalia; Euthera; Chordata; Vertebrata; Euteleostomi;
            Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 128851)
AUTHORS    Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE      Rattus norvegicus, clone RP31-256A2
JOURNAL    Unpublished
REFERENCE   2 (bases 1 to 128851)
AUTHORS    Dewart,K., Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A.,
            Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T.,
            Boguslavskiy,L., Boukhgalter,B., Brown,A., Camarata,J.,
            Gargyala,S., Gordon,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliiev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kellis,C., Lander,E., Allen,A., Linton,L.,
            Mammalia; Euthera; Chordata; Vertebrata; Euteleostomi;
            Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.

Submitted (16-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 128851)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gordon,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliiev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kellis,C., Lander,E., Allen,A., Linton,L.,
Mammalia; Euthera; Chordata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
Submitted (16-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 16, 2002 this sequence version replaced gi:22123285.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

```

Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (01-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 112570)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE  
JOURNAL  
COMMENT  
Submitted (02-MAY-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence.submissions@genome.wi.mit.edu](mailto:sequence.submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L3184  
Center clone name: 256\_A\_2  
-----

FEATURES  
Source  
Location/Qualifiers  
1. .112570  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="RP31-256A2"  
/clone\_lib="RP31-31 Female Rat PAC"  
480. .527  
repeat\_region  
/rpt\_family="(TG)n"  
2529. .2566  
repeat\_region  
/rpt\_family="AT\_rich"  
4028. .4050  
repeat\_region  
/rpt\_family="GC\_rich"

repeat\_region  
/rpt\_family="CT-rich"  
10372. .10451  
repeat\_region  
/rpt\_family="(CCG)n"  
12554. .12595  
repeat\_region  
/rpt\_family="(TTTG)n"  
13284. .13307  
repeat\_region  
/rpt\_family="(TTTG)n"  
14249. .14319  
/note="less than 30 qual SNGL region"  
complement(14249. .14271)  
/note="<30 quality SNGL region"  
complement(14278. .14288)  
/note="<30 quality SNGL region"  
complement(14308. .14315)  
/note="<30 quality SNGL region"  
15321. .15389  
repeat\_region  
/rpt\_family="CT-rich"  
15491. .15534  
repeat\_region  
/rpt\_family="(TTTC)n"  
16602. .16634  
repeat\_region  
/rpt\_family="AT\_rich"  
16639. .16690  
repeat\_region  
/rpt\_family="(ATG)n"  
17526. .17595  
repeat\_region  
/rpt\_family="GA-rich"  
18502. .18556  
repeat\_region  
/rpt\_family="(TG)n"  
19757. .19765  
/note="<30 quality SNGL region"  
20153. .20204  
repeat\_region  
/rpt\_family="(GA)n"  
20250. .20369  
repeat\_region  
/rpt\_family="(GAAA)n"  
21159. .21182  
repeat\_region  
/rpt\_family="AT\_rich"  
23886. .23957  
repeat\_region  
/rpt\_family="(GA)n"  
27479. .27500  
repeat\_region  
/rpt\_family="GC\_rich"  
28430. .28560  
repeat\_region  
/rpt\_family="CT-rich"  
29380. .29424  
repeat\_region  
/rpt\_family="(GGA)n"  
29526. .29558  
repeat\_region  
/rpt\_family="A-rich"  
30184. .30240  
repeat\_region  
/rpt\_family="(T)n"  
36278. .36303  
repeat\_region  
/rpt\_family="GC\_rich"  
37035. .37151  
repeat\_region  
/rpt\_family="LIMDa"  
37173. .37238  
repeat\_region  
/rpt\_family="(GGA)n"  
37626. .37715  
repeat\_region  
/rpt\_family="(TC)n"  
39025. .39079  
repeat\_region  
/rpt\_family="(CA)n"  
42932. .42965  
repeat\_region  
/rpt\_family="(A)n"  
43371. .43392  
repeat\_region  
/rpt\_family="AT\_rich"  
47172. .47236  
repeat\_region  
/rpt\_family="(GA)n"  
53364. .53538  
repeat\_region  
/rpt\_family="G-rich"  
54119. .54167  
repeat\_region  
/rpt\_family="GC\_rich"  
54515. .54583  
repeat\_region  
/rpt\_family="CT-rich"  
55752. .55792  
repeat\_region  
/rpt\_family="(TG)n"  
56475. .56543  
repeat\_region



by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

----- Project Information

Center project name: GIAY

Center clone name: CH230-140P4

----- Summary Statistics

Assembly program: Atlas 3.0;

Consensus quality: 226791 bases at least Q40

Consensus quality: 228171 bases at least Q30

Consensus quality: 229344 bases at least Q20

Estimated insert size: 240235; sum-of-contigs estimation

Estimated insert size: 274125; agarose-fp estimation

Quality coverage: 5x in Q20 bases; agarose-fp estimation

Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 13 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence.  
\* as soon as it is available and the accession number will  
\* be preserved.

```
1 1005: contig of 1005 bp in length
* 1006 3091: gap of 2086 bp
* 3092 4962: contig of 1871 bp in length
* 4963 6209: gap of 1247 bp
* 6210 7864: contig of 1655 bp in length
* 7865 8063: gap of 199 bp
* 8064 18114: contig of 10051 bp in length
* 18115 18226: gap of 112 bp
* 18227 219701: contig of 201475 bp in length
* 219702 219801: gap of unknown length
* 219802 221361: contig of 1560 bp in length
* 221362 221551: gap of 190 bp
* 221552 225319: contig of 3768 bp in length
* 225320 225419: gap of unknown length
* 225420 226560: contig of 1141 bp in length
* 226561 226660: gap of unknown length
* 226661 228063: contig of 1403 bp in length
* 228064 228463: gap of unknown length
* 228464 229384: contig of 1221 bp in length
* 229385 229484: gap of unknown length
* 229485 230873: contig of 1389 bp in length
* 230874 230973: gap of unknown length
* 230974 232227: contig of 1254 bp in length
* 232228 233277: gap of unknown length
* 233278 234603: contig of 2276 bp in length.
```

## FEATURES

source

```
1. .234603
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-140P4"
```

misc\_feature

```
1. .1005
/notes="wgs end_extension"
```

clone end:T7"

misc\_feature

```
3092..4962
/notes="wgs end_extension"
```

clone end:T7"

misc\_feature

```
6210..7864
/notes="wgs end_extension"
```

```
misc_feature
clone_end:T7"
complement(8503..9238)
/notes="clone_boundary"
clone_end:T7"
site:ECORI
end_sequence:BZ094084"
219802..221361
/notes="wgs_contig"

misc_feature
ORIGIN
Query Match      81.6%; Score 199.2; DB 2; Length 234603;
Best Local Similarity 88.5%; Pred. No. 2.6e-47;
Matches 216; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 1 ACGCCGGGGATTAAATGCCACTAAAGCGGTGATCCATCACTCGGAGCGCGCAACT 60
DB 25386 AGCGCGGGGGATTAAATGCCACTAAAGCGGTGATCCATCACTCGGAGCGCGTAAACT 25445
QY 61 TTTCGAGGAGGCTCAGCCATTGGCTGACACCGCTCAGTGCCTCTCTCAGCGTCTCTCG 120
DB 25446 TTCGAGGCGGCTCAGCCATTGGCTGACAGAGTCACTGTCTCTCCCTCGAGCCCTCG 25505
QY 121 CCTCTCCGCCCCCTCTTGGCACTGTACATTCATATATATTTCTTCTCGGCCCCAT 180
DB 25506 CCTGCGCGCGCGCTCTTGGCACTGTACATTCATATATTTCTTCTCGTCTCTAT 25565
QY 181 GGAGAGTGAAGAGTGGGACACTCAGCGCGGCTTCGAGGACGAGTCACTCAGTG 240
DB 25566 GGAGAGTGAAGAGTGGGACCGGTCAACCCAGGCTTCGAGGACCGGCTCACTCAGTG 25625
QY 241 ACAG 244
DB 25626 ACAG 25629
```

## RESULT 7

AC122669

LOCUS

DEFINITION

AC122669

AC122669

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 298853)

Muzny,D,Marie,, Metzker,M, Lee,, Abramson,S,, Adams,C., Alder,J.,

Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,

Anyalebechi,V., Ayodeji,A., Ayodeji,M., Baca,E., Baden,H.,

Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,

Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,

Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,

Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,

Chacko,J., Chavez,D., Chen,R., Chen,Y., Chen,Z., Chu,J.,

Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,

Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,

Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,

Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,

Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,

Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,

Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,

Gebregiorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,

Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,

Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,

Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,

Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlesbird,D., Jackson,A.,

Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,

Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,

Kowis,C., Kraft,C.L., Lebow,H., Levant,J., Lewis,H., Li,Z., Liu,J.,

Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,

Mammaliya; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

AC122669

AC122669

AC122669

AC122669

AC122669





Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Testfaye,S.,  
Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

## Direct Submission

Submitted (29-MAR-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 236214)

Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,  
Anderson,M., Archachchi,H.M., Barna,N., Bastien,V., Bloom,T.,  
Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,  
Collimore,A., Cooke,P., Cooke,P., Corum,B., DeArellano,K.,  
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,  
Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,  
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,  
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kellis,C., Landers,T., Levine,R.,  
Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,  
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,  
Meldrim,J., Meneus,L., Mihova,T., Mieng,V., Murphy,T., Naylor,J.,  
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,  
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,  
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,R., Rogov,P.,  
Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C.,  
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,  
Talamas,J., Testfaye,S., Theodore,J., Topham,K., Travers,M.,  
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,  
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

## Direct Submission

Submitted (17-DEC-2003) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

On Dec 17, 2003 this sequence version replaced gi:31442525.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WITBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L12520

Center clone name: 103\_L\_13

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 9 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 13453: contig of 13453 bp in length  
\* 13454 13553: gap of 100 bp  
\* 13554 24211: contig of 10658 bp in length  
\* 24211 24311: gap of 100 bp  
\* 24311 29735: contig of 5424 bp in length  
\* 29735 29736: gap of 100 bp  
\* 29736 29836: gap of 100 bp  
\* 29836 44484: contig of 14648 bp in length  
\* 44484 44583: gap of 100 bp  
\* 44583 65367: contig of 20784 bp in length  
\* 65367 65468: gap of 100 bp  
\* 65468 75780: contig of 10313 bp in length  
\* 75780 75781: gap of 100 bp  
\* 75781 143236: contig of 67356 bp in length  
\* 143236 143237: gap of 100 bp  
\* 143237 177855: contig of 34519 bp in length  
\* 177855 177856: gap of 100 bp  
\* 177856 177956: contig of 58259 bp in length.

## FEATURES

## source

1. .236214  
/organism="Mus musculus"  
/mol\_type="genomic DNA"

/db\_xref="taxon:10090"  
/chromosome="6"  
/map="6"  
/clones="RP23-103L13"  
/clone\_lib="RPC1-23 Female Mouse BAC"

## ORIGIN

Query Match 75.4%; Score 184; DB 2; Length 236214;  
Best Local Similarity 87.3%; Pred. No. 7e+43;  
Matches 213; Conservative 0; Mismatches 30; Indels 1; Gaps 1;  
QY 1 AGCCCGGGGATTAAATGCCACTAAACGGTGTATCCATCTCGGAGCGCGCAACT 60  
Db 31927 AGCCCGGGGATTAAATGCCACTAAACGGTGTATCCATCTCGGAGCGCGTAAACT 31986  
QY 61 TTTCAGGAGGCTCAGGCATTGGCTGACACGGTCACGTCGCCCTCCCTCCACGCTCCCTCG 120  
Db 31987 TTGCAGGCGGCTCAGGCATTGGCTGAGTGTAGTACAGTG-TCTTCCCTCGAGCGCTCCG 32045  
QY 121 CCTTCCGCGCCCTCTTGGGCACTGTACATTATATATATATATATATATATATATATATATAT 180  
Db 32046 CCTTCCGCGCCCTCTTGGGCACTGTACATTATATATATATATATATATATATATATATAT 32105  
QY 181 GGAGGAAGTGAAGAAGTTGGCAGTCAACCGCGGCTTCCGAGCAGGTCATCTCAGTG 240  
Db 32106 GGAGGAAGTGAAGAAGTTGGCAGTCAACCGCGGCTTCCGAGGATCCAAATCACTCAGTG 32165  
QY 241 ACAG 244  
Db 32166 ACAG 32169

## RESULT 10

## AC123672

## LOCUS

AC123672 236829 bp DNA linear HTG 16-DEC-2003  
DEFINITION Mus musculus chromosome 6 clone RP23-253E11 map 6, \*\*\* SEQUENCING  
IN PROGRESS \*\*\*, 9 unordered pieces.

## AC123672

AC123672.4 GI:39930763

HTG; HTGS PHASE1; HTGS FULLTOP; HTGS\_ACTIVEFIN.

Mus musculus (house mouse)

## SOURCE

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

Birren,B., Nusbaum,C. and Lander,E.

1 (bases 1 to 236829)

Mus musculus chromosome 6, clone RP23-253E11

Unpublished

2 (bases 1 to 236829)

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

TITLE  
JOURNAL  
Submitted (01-JUN-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 236829)  
Birtten,B., Nusbbaum,C., Lander,E., Abouelleil,A., Allen,N.,  
Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,  
Bognuslavsky,L., Boukghalter,B., Camarata,J., Chang,J., Choepel,Y.,  
Collamore,A., Cook,A., Cooke,P., Corium,B., DeAtrellano,K.,  
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,  
Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,  
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,  
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,  
Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,  
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,  
Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,  
Nguyen,C., Nicol,R., Norbu,C., O'Connor,R., O'Donnell,P.,  
O'Neill,D., Oliver,J., Petersen,K., Phunkhang,P., Pierre,N.,  
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,  
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,  
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travets,M.,  
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Willson,B., Wu,X.,  
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (16-DEC-2003) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Dec 16, 2003 this sequence version replaced gi:28975939.  
All repeats were identified using RepeatMasker:  
Smith, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RW/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L26388  
Center clone name: 253\_E\_11  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 9 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
\* 1 13701: contig of 13701 bp in length  
\* 13702 13801: gap of 100 bp  
\* 13802 24723: contig of 10922 bp in length  
\* 24724 24824: gap of 100 bp  
\* 24824 45150: contig of 20327 bp in length  
\* 45151 45250: gap of 100 bp  
\* 45251 66059: contig of 20809 bp in length  
\* 66060 66159: gap of 100 bp  
\* 66160 76371: contig of 10212 bp in length  
\* 76372 76471: gap of 100 bp  
\* 76472 144414: contig of 67943 bp in length  
\* 144415 144514: gap of 100 bp  
\* 144515 169981: contig of 25467 bp in length  
\* 169982 170081: gap of 100 bp  
\* 170082 178409: contig of 8328 bp in length  
\* 178410 178509: gap of 100 bp  
\* 178510 236829: contig of 58320 bp in length.  
Location/Qualifiers  
1. .236829  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="6"  
/map="6"  
/clone="RP23-253E11"  
/clone\_lib="RPCI-23 Female Mouse BAC"  
-----  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT





```
misc_feature      759..920
/Note="Homeobox; Region: Homeobox domain"
/db_xref="CDD:pfam00046"

ORIGIN
Query Match      25.4%; Score 62; DB 9; Length 2529;
Best Local Similarity 100.0%; Pred. No. 3.4e-07; Indels 0; Gaps 0;
Matches 62; Conservative 0; Mismatches 0;

QY 183 AGGAAGTGAAGAAGTTGGCAGTCACGCCGGCTTCGAGGACCAAGTCACTCAGTAC 242
    |||||
Db 1 AGGAAGTGAAGAAGTTGGCAGTCACGCCGGCTTCGAGGACCAAGTCACTCAGTAC 60

QY 243 AG 244
    ||
Db 61 AG 62

RESULT 14
MUSRA1A          2216 bp mRNA linear ROD 12-JUN-1993
LOCUS            Mouse ERA-1-993 mRNA, complete cds, and alternate ERA-1-339 mRNA,
DEFINITION       complete cds.
ACCESSION        M22115
VERSION          M22115.1 GI:193047
KEYWORDS         ERA-1 protein, early retinoic acid protein.
SOURCE           Mus musculus
ORGANISM         Mus musculus (house mouse)
REFERENCE        1 (bases 1 to 2216)
AUTHORS          LaRosa,G.J. and Gudas,L.J.
TITLE            Early retinoic acid-induced F9 teratocarcinoma stem cell gene
                ERA-1: alternate splicing creates transcripts for a
                homeobox-containing protein and one lacking the homeobox
JOURNAL          Mol. Cell. Biol. 8 (9), 3906-3917 (1988)
MEDLINE          89127233
PUBMED           2906112
COMMENT          Original source text: Mouse (F9 teratocarcinoma stem cell line)
                mRNA, clone ERA-1-993.
FEATURES         Location/Qualifiers
source           1..2216
                /organism="Mus musculus"
                /mol_type="mRNA"
                /db_xref="taxon:10090"
                81..1076
                /note="ERA-1-993 protein"
                /codon_start=1
                /db_xref="GI:309218"
                /translation="MNSFLEYPILGSDGTCARAYPSDHGITTFCSCAVSANS CGG
                DDFLVGRGVQISSLPHHHHHHHHPQTATYQTSNGLGISYSHS SCGFSYAQNPSA
                PYGPTGLNQADVSGYPPCAPVYGNLSTPMVQHSHHGHQYAGTGVSPQYIHHSY
                GQEQTLALATYNNLSPLHAGHQACRPASETS PAQTFDMVKRNPPTKRGVE
                YGVGPNVARNFTTKLTLEKEFHFNKYLTRARRVEIAASLOLNQVQKWFQNR
                RMKQKREKGLLPISBPATPPGSDGTEESSEKSPSPSPASPSTDTLTTH"
                join(81..422,626..685)
                /note="ERA-1-339 protein"
                /codon_start=1
                /db_xref="GI:387146"
                /translation="MNSFLEYPILGSDGTCARAYPSDHGITTFCSCAVSANS CGG
                DDFLVGRGVQISSLPHHHHHHHHPQTATYQTSNGLGISYSHS SCGFSYAQNPSA
                PYGPTGLNQADVPPRSLSPCFPRDVFSSADL"
                <81..422
                /note="ERA-1-399 protein"
                626..685
                /note="ERA-1-399 protein"
                753..932
                /note="ERA-1-993 protein homeobox"

misc_feature      753..932
/Note="ERA-1-993 protein homeobox"

ORIGIN
Query Match      22.0%; Score 53.6; DB 10; Length 2216;
Matches 58; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 180 TGGAGGAAGTGAAGAAGTTGGCAGTCACGCCGGCTTCGAGGACCAAGTCACTCAGT 239
    |||||
Db 1 TGGAGGAAGTGAAGAAGTTGGCAGTCACGCCGGCTTCGAGGACCAAGTCACTCAGT 60

QY 240 GACAG 244
    |||||
Db 61 GACAG 65

RESULT 15
MUSHOX161        443 bp DNA linear ROD 01-NOV-2000
LOCUS            Mus musculus Hox1.6 protein (Hox1.6) gene, exon 2.
DEFINITION       M20214 M15928
ACCESSION        M20214.1 GI:193918
VERSION          M20214.1
KEYWORDS         homeo box.
SEGMENT          1 of 3
SOURCE           Mus musculus (house mouse)
ORGANISM         Mus musculus
REFERENCE        1 (bases 1 to 443)
AUTHORS          Baron,A., Featherstone,M.S., Hill,R.E., Hall,A., Galliot,B. and
                Duboule,D.
TITLE            Hox-1.6: a mouse homeo-box-containing gene member of the Hox-1
                complex
JOURNAL          EMBO J. 6 (10), 2977-2986 (1987)
MEDLINE          88082669
PUBMED           2891503
FEATURES         Location/Qualifiers
source           1..443
                /organism="Mus musculus"
                /mol_type="genomic DNA"
                /db_xref="taxon:10090"
                /clone_lib="library of H. Lehrach"
                /dev_stage="embryo"
                20..435
                /gene="Hox1.6"
                /number=2

ORIGIN
Query Match      20.9%; Score 51; DB 10; Length 443;
Best Local Similarity 86.1%; Pred. No. 0.0006;
Matches 68; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

QY 166 CTCTCCGGCCCCCATGGAGGAAGTGGCAGAGTGGCAGAGTTCGCGGGCTTCGACGA 225
    |||||
Db 1 CTCTCTGGTCTCTATGGAGGAAGTGGCAGAGTGGCAGAGTTCGCGGGCTTCGACGA 59

QY 226 CCAGGTCACTCAGTGACAG 244
    |||||
Db 60 TCCAATCACTCAGTGACAG 78

Search completed: September 18, 2004, 12:08:43
Job time : 1172.45 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 2004, 09:40:06 ; Search time 262.692 Seconds  
(without alignments)  
3945.922 Million cell updates/sec

Title: US-10-614-282-2

Perfect score: 244

Sequence: 1 agcgccggggatttaaatgc.....accaggctcaactcagtgacag 244

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

N Geneseq 29Jan04:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002s:\*  
7: geneseqn2003as:\*  
8: geneseqn2003bs:\*  
9: geneseqn2003cs:\*  
10: geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	172.6	70.7	6118	6 ABN80111	ABN80111 Human che
2	135.2	55.4	6118	6 ABN80110	ABN80110 Human che
C 3	38	15.6	302250	6 ABL67703	ABL67703 Oesophagu
4	37.8	15.5	58857	3 AAA58471	AAA58471 Nucleotid
5	37.2	15.2	3050	9 ADB62108	ADB62108 Human cDN
6	37	15.2	38110	7 AAL53730	AAL53730 Ribonucle
7	37	15.2	38110	7 AAD49662	AAD49662 Mycobacte
8	37	15.2	63155	9 ADC85996	ADC85996 Human GPC
9	37	15.2	110000	4 AAI99682_25	Continuation (26 o
10	37	15.2	110000	4 AAI99683_25	Continuation (26 o
11	36	14.8	374	6 ABL81906	ABL81906 Human ova
12	35.6	14.6	1141	9 ADC87092	ADC87092 Human GPC
13	35.6	14.6	2179	7 ADA53147	ADA53147 Human cDN
14	35.6	14.6	2818	9 ADB62471	ADB62471 Human cDN
15	35.6	14.6	3089	9 ADB62760	ADB62760 Human cDN
16	35.6	14.6	3743	9 ADB62119	ADB62119 Human cDN
17	35.6	14.6	11706	5 ABA20479	ABA20479 Human ner
18	35.6	14.6	34875	9 ADD14677	ADD14677 Human src
C 19	35.4	14.5	541	9 ADB49520	ADB49520 Primary r
C 20	35	14.3	2363	4 ABL17819	ABL17819 Drosophil
C 21	35	14.3	2876	4 ABL17828	ABL17828 Drosophil
C 22	35	14.3	4363	4 ABL17818	ABL17818 Drosophil
23	34.6	14.2	3060	2 AAX39667	AAX39667 Renal can

24	34.6	14.2	4725	5 AAS76375	Aas76375 DNA encod
25	34.6	14.2	4739	6 ABL63316	Abi63316 Breast ca
26	34.6	14.2	4739	7 ACC00031	Acc00031 Human Bcr
C 27	34.6	14.2	4756	5 AAS85030	Aas85030 DNA encod
28	34.6	14.2	4775	5 AAS76377	Aas76377 DNA encod
29	34.6	14.2	5795	5 AAS85031	Aas85031 DNA encod
30	34.4	14.1	122888	6 ABK83569	ABK83569 Human cDN
C 31	34	13.9	175737	6 ABK83571	ABK83571 Human cDN
32	33.6	13.8	1778	6 AAS18963	Aas18963 Wheat ace
C 33	33.4	13.7	2537	6 ABL51803	Abi51803 Human mtd
C 34	33.4	13.7	2550	9 ADD29657	Add29657 Human tum
C 35	33.4	13.7	16161	6 AAD36071	Aad36071 Human adi
36	33	13.5	152141	7 ACA64961	Aca64961 Human BCR
37	32.8	13.4	33137	9 ADC86460	Adc86460 Human GPC
38	32.8	13.4	48840	8 ACH03435	Ach03435 Genomic D
C 39	32.6	13.4	3337	4 ABL27738	Abi27738 Drosophil
C 40	32.4	13.3	315	4 AAF85371	Aaf85371 Degenerat
C 41	32.4	13.3	315	7 ABX12105	Abx12105 Degenerat
C 42	32.4	13.3	4769	3 AAA50439	Aaa50439 Porcine a
43	32.4	13.3	5349	3 AAA65527	Aaa65527 Porcine B
44	32.4	13.3	5577	8 ADB12722	Adb12722 Human pol
C 45	32.4	13.3	12001	2 AAQ76213	Aaq76213 HSV L/St

## ALIGNMENTS

### RESULT 1

ABN80111/c

ID ABN80111 standard; DNA; 6118 BP.

AC ABN80111;

DT 15-JUL-2002 (first entry)

Human chemically modified disease associated gene SEQ ID NO 128.

Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;  
heart disease; epilepsy; histone deacetylation; muscular dystrophy;  
dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;  
antidiabetic; cytostatic; anticonvulsant; ds.

OS Homo sapiens.

OS Synthetic.

XX WO200200927-A2.

XX 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP007536.

XX 30-JUN-2000; 2000DE-01032529.

XX 01-SEP-2000; 2000DE-01043826.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130908/17.

PT Novel nucleic acid useful for diagnosis and therapy of diseases  
associated with development genes such as diabetes, comprises a sequence  
of a segment of chemically pretreated DNA of genes associated with  
development.

PS Claim 1; SEQ ID NO 128; 27pp; English.

XX The invention relates to a nucleic acid (I) comprising a sequence at  
least 18 bases in length of a segment of chemically pretreated DNA (II)  
of genes associated with development selected from 87 genes listed in the  
specification such as ACCPN, ADFN, or AFDI and comprising one of 350  
sequences (ABN79984-ABN80333) or their complements. The invention is  
useful for the diagnosis or therapy of diseases associated with





FT	CDS	32893..34830		PT	Chemically modifying biological molecules.
FT		/*tag= j		XX	
FT		/note= "ORF 21; encodes AAB07565"		PS	Claim 8; Page 97-136; 162pp; English.
FT	CDS	34827..35804		XX	
FT		/*tag= k		CC	The present sequence represents the BLM (Bleomycin) gene cluster, containing open reading frames (ORFs) 8-30. The proteins encoded by the gene cluster are useful for producing peptides and/or polypeptide
FT		/transl_except= (pos: 1..3, aa: Met)		CC	gene cluster are especially bleomycin or bleomycin analogues. They are also
FT		/note= "ORF 20; encodes AAB07566"		CC	metabolites, especially bleomycin or bleomycin analogues. They are also
FT	CDS	35818..37302		CC	useful for chemically modifying biological molecules to produce branched
FT		/*tag= l		CC	metabolites, especially bleomycin or bleomycin analogues. They are also
FT		/transl_except= (pos: 1..3, aa: Met)		CC	useful for chemically modifying biological molecules to produce branched
FT		/note= "ORF 19; encodes AAB07567"		CC	methyle groups, and for coupling amino acids and fatty acids. They may be
FT	CDS	37299..39215		CC	reacted with an apo-carrier protein and coenzyme A to produce a holo-
FT		/*tag= m		CC	carrier protein. The BLM gene cluster or catalytic domains can be used
FT		/transl_except= (pos: 1..3, aa: Met)		CC	individually or collectively to produce thiazolidine, thiazoline,
FT		/note= "ORF 18; encodes AAB07568"		CC	bithiazoline and bithiazoline-containing microbial metabolites. The BLM
FT	CDS	39301..47181		CC	gene cluster may also be used to produce sugars
FT		/*tag= n		XX	
FT		/note= "ORF 17; encodes AAB07569"		SQ	Sequence 58857 BP; 7256 A; 25139 C; 19353 G; 7109 T; 0 U; 0 Other;
FT		47178..49985		Query Match	15.5%; Score 37.8; DB 3; Length 58857;
FT		/*tag= o		Best Local Similarity	50.8%; Pred. No. 0.88;
FT		/transl_except= (pos: 1..3, aa: Met)		Matches	90; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
FT		/note= "ORF 16; encodes AAB07570"		QY	44 GCGGAAGCGCGCAACTTTTGCAGAGGCTCAGCATTGGTGACACGTCACGTGCCCC 103
FT	CDS	49982..51001		Db	27520 GCGGACCGCGGACCGTGTGGACGGGTGCGACGCGCGCACCCCTTCGGGCGCG 27579
FT		/*tag= p		QY	104 TCCTCAGGTCCTCGCCCTCCGCCCCCTCTTGGCAGTGTACATTATATATATT 163
FT		/transl_except= (pos: 1..3, aa: Met)		Db	27580 TCGTCTCCGCTCTCGCCCCCGCGCACGCCCGCGCGCGCTCTCTGCAGACCTGT 27639
FT		/note= "ORF 15; encodes AAB07571"		QY	164 TTCTTCTCGGCCCCCATGGAGGAAGTGGACAAAGTTGGCACAGTCACGCGGGCTTCG 220
FT		/*tag= q		Db	27640 TCGTCTCAGCGCGAGTACGCGACGAGGCGGACGCGGTACCGCGCTCGCCCTGG 27696
FT		/transl_except= (pos: 1..3, aa: Met)		RESULT 5	
FT		/note= "ORF 14; encodes AAB07572"		ADB62108	
FT	CDS	52383..52946		ID	ADB62108 standard; cDNA; 3050 BP.
FT		/*tag= r		XX	
FT		/transl_except= (pos: 1..3, aa: Met)		AC	ADB62108;
FT		/note= "ORF 13; encodes AAB07573"		DT	
FT	CDS	53018..54130		XX	04-DEC-2003 (first entry)
FT		/*tag= s		XX	Human cDNA encoding clone BRAWH20063010.
FT		/note= "ORF 12; encodes AAB07574"		XX	Human; ss; gene; pharmaceutical; diagnostic; gene therapy;
FT	CDS	54187..55824		KW	tissue regeneration; cell regeneration; membrane protein;
FT		/*tag= t		KW	signal transduction-related protein; transcription-related protein;
FT		/note= "ORF 11; encodes AAB07575"		KW	osteoporosis; neurological disease; cancer; tumour.
FT	CDS	55821..56093		XX	
FT		/*tag= u		OS	Homo sapiens.
FT		/transl_except= (pos: 1..3, aa: Met)		XX	
FT		/note= "ORF 10; encodes AAB07576"		XX	Key
FT	CDS	56090..57586		FT	Location/Qualifiers
FT		/*tag= v		FT	1839..2579
FT		/transl_except= (pos: 1..3, aa: Met)		FT	/*tag= a
FT		/note= "ORF 9; encodes AAB07577"		FT	/product= "Clone BRAWH20063010 protein"
FT	CDS	57583..58857		XX	
FT		/*tag= w		XX	EP1308459-A2.
FT		/note= "ORF 8; encodes AAB07578"		XX	
XX	XX	WC2000040704-A1.		XX	07-MAY-2003.
XX	XX	13-JUL-2000.		XX	28-MAR-2002; 2002EP-00007401.
XX	XX	06-JAN-2000; 2000WO-US000445.		XX	05-NOV-2001; 2001JP-00379298.
XX	XX	06-JAN-1999; 99US-0115435P.		XX	25-JAN-2002; 2002US-00350978.
XX	XX	05-FEB-1999; 99US-0118948P.		XX	(HELI-) HELIX RES INST.
XX	XX	05-JAN-2000; 2000US-00477962.		XX	(REAS-) RES ASSOC BIO TECHNOLOGY.
XX	XX	(REGC ) UNIV CALIFORNIA.		XX	Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
XX	XX	Shen B, Du L, Sanchez C, Chen M, Edwards DJ;		XX	Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
XX	XX	WPI: 2000-465974/40.		XX	Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX	XX	P-PSDB; AAB07556, AAB07557, AAB07558, AAB07559, AAB07560, AAB07561,		XX	
XX	XX	AAB07562, AAB07563, AAB07564, AAB07565, AAB07566, AAB07567, AAB07568,		XX	
XX	XX	AAB07569, AAB07570, AAB07571, AAB07572, AAB07573, AAB07574, AAB07575,		XX	
XX	XX	AAB07576, AAB07577, AAB07578.		XX	
XX	XX	New bleomycin gene cluster components useful for peptide and/or		XX	
XX	XX	polypeptide metabolites, especially bleomycin, production and for		XX	

DR WPI; 2003-450961/43.  
DR P-PSDB; ADB64078.  
XX  
XX  
PT New polynucleotides and polypeptides, useful for developing a diagnostic  
PT marker or medicines for regulation of their expression and activity, or  
PT as targets of gene therapy.  
XX  
XX  
PS Claim 1; Page; 222pp; English.  
XX  
XX  
CC The invention discloses a polynucleotide comprising a sequence selected  
CC from 1970 fully defined nucleotide sequences which encode novel  
CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide  
CC or its partial peptide, an antibody binding to the polypeptide or peptide  
CC of the polynucleotide, immunologically assaying the polypeptide or  
CC peptide of the polynucleotide by contacting the polypeptide or peptide  
CC with the antibody of the encoded protein, and observing the binding  
CC between the two, a transformant carrying the polynucleotide in an  
CC expressible manner and an antisense polynucleotide. The oligonucleotide  
CC is useful as a primer for synthesising the polynucleotide, or as a probe  
CC for detecting the polynucleotide. The polynucleotides and encoded  
CC proteins are useful as pharmaceutical agents and many disease-related  
CC genes may be included in them, for developing a diagnostic marker or  
CC medicines for regulation of their expression and activity, or as targets  
CC of gene therapy. The genes are involved in tissue and/or cell  
CC regeneration. Membrane proteins, signal transduction-related proteins,  
CC transcription-related proteins, disease-related proteins and genes  
CC encoding them can be used as indicators for diseases (e.g. osteoporosis,  
CC neurological diseases, cancer, tumours). The cDNA may be used to regulate  
CC the activity or expression of the encoded protein to treat diseases. The  
CC sequence presented is a cDNA of the invention. Note: Some of the sequence  
CC data for this patent is not represented in the printed specification, but  
CC is based on sequence information supplied by the European Patent Office.  
XX  
XX  
SQ Sequence 3050 BP; 491 A; 1023 C; 953 G; 583 T; 0 U; 0 Other;  
  
Query Match 15.2%; Score 37.2; DB 9; Length 3050;  
Best Local Similarity 56.6%; Pred. No. 0.52;  
Matches 69; Conservative 0; Mismatches 53; Indels 0; Gaps 0;  
  
QY 29 CGGTGATCCATCATCTGGAGCGCGCAACTTTGCGAGGAGGCTCAGCCATTGGCTGAC 88  
DB 2027 CGTCTCCCTCCCTCGGAGCCGCGCTCGGCCCTCCCTCGGAGCCGCTCCCTCC 2086  
  
QY 89 ACCGTACGTGCCCTCTCTCCAGCGTCTCTCGCCCTCCCGCCCCCTCTTGGCACTGT 148  
DB 2087 AGCGACGCGCTCTCTCCCTCCCTGCGGACCTCGTCCCTCCCTCCCTGCGGACTGT 2146  
  
QY 149 AC 150  
DB 2147 GC 2148  
  
RESULT 6  
AAL53730  
ID AAL53730 standard; DNA; 38110 BP.  
XX  
XX  
AC AAL53730;  
AC  
DT 07-FEB-2003 (first entry)  
XX  
DE Ribonuclease P RNAse P DNA SEQ ID No 24.  
XX  
XX Target RNA; target RNA; support-attached test compound; flow cytometry;  
XX mass spectrometry; high-throughput screening; ds.  
XX  
XX Mycobacterium tuberculosis.  
XX  
XX WO200283837-A1.  
XX  
XX 24-OCT-2002.  
XX  
XX 11-APR-2002; 2002WO-US011758.  
XX  
XX (PTCT-) PTC THERAPEUTICS INC.

PR 11-APR-2001; 2001US-0282966P.  
XX  
XX (PTCT-) PTC THERAPEUTICS INC.  
XX  
XX Almstead NG;  
XX  
XX WPI; 2003-075534/07.  
XX  
XX  
PT Identifying a test compound that binds to a target RNA molecule by  
PT separating the detectably labeled target RNA: support-attached test  
PT compound complex from uncomplexed target RNA molecules and test compounds  
XX by flow cytometry.  
XX  
XX Example; Page 61-79; 131pp; English.  
XX  
XX The invention relates to a novel method for identifying a test compound  
XX that binds to a target RNA molecule comprising separating the detectably  
XX labeled target RNA: support-attached test compound complex from  
XX uncomplexed target RNA molecules and test compounds. The separating  
XX process is carried out by flow cytometry and determining a structure of  
XX the type of test compound of the RNA: support-attached test compound  
XX complex by mass spectrometry. The method is useful for high-throughput  
XX screening of libraries of compounds to identify pharmaceutical leads.  
XX This polynucleotide sequence represents a DNA sequence related to the  
XX detecting method of the invention  
XX  
SQ Sequence 38110 BP; 6702 A; 12862 C; 12361 G; 6185 T; 0 U; 0 Other;  
  
Query Match 15.2%; Score 37; DB 7; Length 38110;  
Best Local Similarity 52.2%; Pred. No. 1.3;  
Matches 82; Conservative 0; Mismatches 75; Indels 0; Gaps 0;  
  
QY 45 CGGAAGCGCGCAACTTTTGCAGGAGGCTCAGCCATTGGCTGACACCGTACGTGCCCT 104  
DB 35959 CGGTGGCGCGCTACATGTGGCGGCGGAGGAGGTTGGCGCTCGGCCCGGTACCGG 36018  
  
QY 105 CTTCCAGGTCTCCGCCCTCCCGCCCCCTCTTGGCACTGTACATTATCATTTT 164  
DB 36019 CCACCAGCACTCGGGCAATCCGGCGCCGATCCGGGGCGGCCAGGGTTCGCGGTG 36078  
  
QY 165 TCTTCTCGGCCCCCATCGAGGAGTGTGAGAAAGTTGGC 201  
DB 36079 CGCGCATCAGCGGAGGTTGGAGTGGCGCACTTCGC 36115  
  
RESULT 7  
AAD49662  
ID AAD49662 standard; DNA; 38110 BP.  
XX  
XX  
AC AAD49662;  
XX  
XX 24-MAR-2003 (first entry)  
XX  
XX Mycobacterium tuberculosis ribonuclease P (RNAse) rnpB DNA.  
XX  
XX  
XX Amyloidosis; haemophilia; Alzheimer's disease; atherosclerosis; cancer;  
XX gigantism; dwarfism; hypothyroidism; hyperthyroidism; cystic fibrosis;  
XX autoimmune disorder; aging; inflammation; diabetes; obesity; anorectic;  
XX neurodegenerative disorder; Parkinson's disease; gene therapy; virucide;  
XX haemostatic; antibacterial; nootropic; neuroprotective; cytostatic;  
XX fungicide; ribonuclease P; RNAse; enzyme; gene; ds.  
XX  
XX Mycobacterium tuberculosis.  
XX  
XX WO200283953-A1.  
XX  
XX 24-OCT-2002.  
XX  
XX 11-APR-2002; 2002WO-US011757.  
XX  
XX 11-APR-2001; 2001US-0282965P.  
XX  
XX (PTCT-) PTC THERAPEUTICS INC.

XX Rando R, Welch E;  
XX WPI; 2003-075561/07.  
XX Identifying a test compound that binds to a target RNA molecule for  
PT treating or preventing amyloidosis, hemophilia, cancer, gigantism,  
PT diabetes, by contacting a detectably labeled target RNA molecule with a  
PT library of test compounds.  
XX Example; Page 70-88; 152pp; English.  
XX The invention relates to a method for identifying a test compound that  
CC binds to a target RNA molecule, which comprises contacting a detectably  
CC labelled target RNA molecule with a library of test compounds under  
CC conditions that permit direct binding of the labelled target RNA to a  
CC member of the library of test compounds so that a detectably labeled  
CC target RNA: test compound complex is formed. The method is useful for  
CC screening libraries of compounds for those that are selectively bind to a  
CC pre-selected target RNA. The compounds are useful for inhibiting the  
CC formation of a specific bound RNA: host cell factor complexes in vivo.  
CC They are also useful for treating or preventing diseases associated with  
CC overproduction or decreased protein function, such as amyloidosis,  
CC haemophilia, Alzheimer's disease, atherosclerosis, cancer, gigantism,  
CC dwarfism, hypothyroidism, hyperthyroidism, autoimmune disorders, aging,  
CC inflammation, cystic fibrosis, diabetes, obesity, neurodegenerative  
CC disorders, Parkinson's disease or infections (bacterial, viral, fungal).  
CC The invention is also used in gene therapy. The present sequence is  
CC Mycobacterium tuberculosis ribonuclease P (RNAase) rnpB DNA. This  
CC sequence is used to illustrate the method of the invention  
XX  
SQ Sequence 38110 BP; 6702 A; 12862 C; 12361 G; 6185 T; 0 U; 0 Other;  
Query Match 15.2%; Score 37; DB 7; Length 38110;  
Best Local Similarity 52.2%; Pred. No. 1.3;  
Matches 82; Conservative 0; Mismatches 75; Indels 0; Gaps 0;  
QY 45 CGGAGCGGCGGCTTTGCGAGGAGTTCAGCCATGCGTACACCGTCAGTGCCT 104  
Db 35959 CGGTGGCGGCTCACATGTGGCGGCGGAGCGAGCTGGCGGCTCGGCCCGTACCGG 36018  
QY 105 CTCCAGGCTCTCCGCGCTCCGCGCCCGCTCTTGGCAGCTGATCATTCATCATTTT 164  
Db 36019 CACCGAGGAGTCTGGGCAATCCGCGCCGATCCGGGGGCGCCAGGGTTCGCGG 36078  
QY 165 TCTTCTCCGCGCCCATGGAGGAAGTGGAGAAAGTTGGC 201  
Db 36079 CGCCGATCAGCGGAGGTTGCGAGTGGCGGACTTCGC 36115  
RESULT 8  
ADC85996  
ID ADC85996 standard; DNA; 63155 BP.  
AC ADC85996;  
XX 01-JAN-2004 (first entry)  
DT  
XX Human GPCR gene SEQ ID NO:449.  
DE  
XX ds; gene; human; GPCR;  
KW guanosine triphosphate-binding protein coupled receptor; gene therapy.  
XX Homo sapiens.  
OS  
XX EP1270724-A2.  
PN  
XX 02-JAN-2003.  
PD  
XX 18-JUN-2002; 2002EP-00013517.  
PF  
XX 18-JUN-2001; 2001JP-00246789.  
PR  
XX

PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.  
PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.  
PI Suwa M, Asai K, Akiyama Y, Aburatani H;  
XX WPI; 2003-315783/31.  
DR P-PSDB; ADC85997.  
DR  
XX New polynucleotide, useful for preparing a composition for treating a  
PT patient in need of increased or suppressed activity or expression of the  
PT guanosine triphosphate-binding protein coupled receptor.  
XX  
PS Claim 1; SEQ ID NO 449; 28pp; English.  
XX  
CC The invention relates to a novel polynucleotide encoding a guanosine  
CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of  
CC the invention may have a use in gene therapy. The polynucleotide and  
CC polypeptide are useful for preparing a composition for treating a patient  
CC in need of increased or suppressed activity or expression of the  
CC guanosine triphosphate-binding protein coupled receptor. The  
CC polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the  
CC invention.  
XX  
SQ Sequence 63155 BP; 18875 A; 12104 C; 12350 G; 18904 T; 0 U; 922 Other;  
Query Match 15.2%; Score 37; DB 9; Length 63155;  
Best Local Similarity 64.7%; Pred. No. 1.6;  
Matches 55; Conservative 0; Mismatches 30; Indels 0; Gaps 0;  
QY 96 CGTGGCCCTCTCCAGCGCTCTCCGCGCTCCGCGCCCGCTCTTGGGCACTGATTCATCA 155  
Db 60851 CTTCT 60910  
QY 156 TATCATTTTCT 180  
Db 60911 TCT 60935  
RESULT 9  
AA199682\_25  
Continuation (26 of 45) of AA199682 from base 2500001 (Mycobacterium tuberculosis strata  
WP Sequence split into 45 fragments LOCUS AA199682 Accession AA199682  
WP Fragment Name Begin End  
WP AA199682\_00 1 110000  
WP AA199682\_01 100001 210000  
WP AA199682\_02 200001 310000  
WP AA199682\_03 300001 410000  
WP AA199682\_04 400001 510000  
WP AA199682\_05 500001 610000  
WP AA199682\_06 600001 710000  
WP AA199682\_07 700001 810000  
WP AA199682\_08 800001 910000  
WP AA199682\_09 900001 1010000  
WP AA199682\_10 1000001 1110000  
WP AA199682\_11 1100001 1210000  
WP AA199682\_12 1200001 1310000  
WP AA199682\_13 1300001 1410000  
WP AA199682\_14 1400001 1510000  
WP AA199682\_15 1500001 1610000  
WP AA199682\_16 1600001 1710000  
WP AA199682\_17 1700001 1810000  
WP AA199682\_18 1800001 1910000  
WP AA199682\_19 1900001 2010000  
WP AA199682\_20 2000001 2110000  
WP AA199682\_21 2100001 2210000  
WP AA199682\_22 2200001 2310000  
WP AA199682\_23 2300001 2410000  
WP AA199682\_24 2400001 2510000  
WP AA199682\_25 2500001 2610000  
WP AA199682\_26 2600001 2710000  
WP AA199682\_27 2700001 2810000  
WP AA199682\_28 2800001 2910000  
WP AA199682\_29 2900001 3010000



[illegible]

RESULT 10					
AAI99683_25					
Continuation (26 of 44) of AAI99683 from base 2500001 (Mycobacterium tuberculosis strain					
WP	WP	WP	Sequence split into 44 fragments	LOCUS AAI99683 Accession AAI99683	
WP	WP	Fragment Name	Begin	End	
WP	AAI99683_00		1	110000	
WP	AAI99683_01		100001	210000	
WP	AAI99683_02		200001	310000	
WP	AAI99683_03		300001	410000	
WP	AAI99683_04		400001	510000	
WP	AAI99683_05		500001	610000	
WP	AAI99683_06		600001	710000	
WP	AAI99683_07		700001	810000	
WP	AAI99683_08		800001	910000	
WP	AAI99683_09		900001	1010000	
WP	AAI99683_10		1000001	1110000	
WP	AAI99683_11		1100001	1210000	
WP	AAI99683_12		1200001	1310000	
WP	AAI99683_13		1300001	1410000	
WP	AAI99683_14		1400001	1510000	
WP	AAI99683_15		1500001	1610000	
WP	AAI99683_16		1600001	1710000	
WP	AAI99683_17		1700001	1810000	
WP	AAI99683_18		1800001	1910000	
WP	AAI99683_19		1900001	2010000	
WP	AAI99683_20		2000001	2110000	
WP	AAI99683_21		2100001	2210000	
WP	AAI99683_22		2200001	2310000	
WP	AAI99683_23		2300001	2410000	
WP	AAI99683_24		2400001	2510000	
WP	AAI99683_25		2500001	2610000	
WP	AAI99683_26		2600001	2710000	
WP	AAI99683_27		2700001	2810000	
WP	AAI99683_28		2800001	2910000	
WP	AAI99683_29		2900001	3010000	
WP	AAI99683_30		3000001	3110000	
WP	AAI99683_31		3100001	3210000	
WP	AAI99683_32		3200001	3310000	
WP	AAI99683_33		3300001	3410000	
WP	AAI99683_34		3400001	3510000	

WP	AAI99683_35	3500001	3610000
WP	AAI99683_36	3600001	3710000
WP	AAI99683_37	3700001	3810000
WP	AAI99683_38	3800001	3910000
WP	AAI99683_39	3900001	4010000
WP	AAI99683_40	4000001	4110000
WP	AAI99683_41	4100001	4210000
WP	AAI99683_42	4200001	4310000
WP	AAI99683_43	4300001	4403765

Query Match            15.2%;    Score 37; DB 4; Length 110000;  
Best Local Similarity   52.2%;    Pred. No. 1.9;  
Matches     82; Conservative     0; Mismatches     75; Indels     0; Gaps     0;

Qy	45	CGGAAGCCGCCAATTTTGCAGGAGCCTCAGCCATTGGCTCACACCGTCACGTGCCCT	104
Db	20637	CGGTGGSCGCTCACATGTGGCGGGCAGCGACGTTGGCCGCTTGCGCCCGTACC	106
Qy	105	CCTTCGAGCTCTCCGCGCTCCGCCCCCTCTTGGCGACGTGATCATATCATTTT	164
Db	20697	CCACGACGACTCGGCAATTCGGCGCCGATCCGGGGCCGCCACAGGTTGCGCGTG	20756
Qy	165	TCTTCTCGGCCCATCGAGGAAGTGAGAACTTGC	201
Db	20757	CGCGATCAGCGCAGGTTCGAGTGGCACTTCGC	20793

RESULT 11  
ABL81906  
ID ABL81906 standard; cDNA; 374 BP.  
XX  
AC ABL81906;  
XX  
DT 17-MAY-2002 (first entry)

RESULT 11  
APT. 81906

ABL81906  
ID ABL81906 standard; cDNA; 374 BP.

AC ABL81906;

DT 17-MAY-2002 (first entry)

Human ovarian cancer related cDNA clone SEO ID NO:4884.

KW Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.

OS Homo sapiens.

PN WO200192581-A2.

PD 06-DEC-2001.

29-MAY-2001; 2001WO-US017756.

PR 26-MAY-2000; 2000US-0207484P.

PA (CORI-) CORIXA CORP.

PI Algate PA, Harlocker SL, Jones R:

DR WPI; 2002-122075/16.

Composition for therapy and diagnosis of ovarian cancer comprising polypeptide of a ovarian tumor polypeptide, polynucleotide encoding polypeptide, antibody specific to polypeptide or T cell expressing polypeptide.

PS Claim 1; SEQ ID NO 4884; 489pp; English.

The present invention describes a composition (I) comprising: carriers and immunostimulants; and a polypeptide (II) of an ovarian tumour polypeptide encoded by a polynucleotide (III) having a cDNA sequence (S1) from the 10912 nucleotide sequences as given in ABL77023 to ASU87934, (III) encoding (II) having a sequence (S2), a T cell population of (II), or antigen presenting cells that express (II). (I) has cytostatic activity. An oligonucleotide (IV) that hybridises to (S1) can be used for detecting ovarian cancer in a patient's biological sample preferably serum or ovarian tissue. The method comprises contacting a biological sample from a patient with (IV), detecting the amount of polynucleotide hybridising to (IV) and comparing the amount to a predetermined cutoff

CC	value and thereby detecting ovarian cancer in the patient, where the	CC	invention.
CC	amount of polynucleotide hybridising to (IV) is detected preferably by	XX	Sequence 1141 BP; 128 A; 440 C; 337 G; 236 T; 0 U; 0 Other;
CC	polymerase chain reaction (PCR). (I) comprising (iii) and/or (ii) is	SQ	Query Match
CC	useful for stimulating and/or expanding T cells specific for an ovarian		14.6%; Score 35.6; DB 9; Length 1141;
CC	tumour protein comprising contacting T cells with (iii) or (ii). (iii) is		Best Local Similarity 55.7%; Pred. No. 1.2;
CC	useful in design and preparation of ribozyme molecules for inhibiting		Matches 68; Conservative 0; Mismatches 54; Indels 0; Gaps 0;
CC	expression of the tumour polypeptides and proteins in tumour cells; and		
CC	to isolate a full length gene from a suitable library e.g., a tumour cDNA		
CC	library using well known techniques		
XX	Sequence 374 BP; 69 A; 101 C; 104 G; 100 T; 0 U; 0 Other;		
SQ	Query Match		
	14.8%; Score 36; DB 6; Length 374;		
	Best Local Similarity 51.9%; Pred. No. 0.61; Indels 0; Gaps 0;		
	Matches 81; Conservative 0; Mismatches 75; Indels 0; Gaps 0;		
QY	10 GATTAAATGCCACTAAACGGTGTCCATCTCGGGAAGCGGCAAACTTTTGCAGGA 69	QY	29 CGGTGATCATCATCTGGCGAGCGGCAAACTTTTGCAGGAGGTCACGCCATTGGCTGAC 88
DB	121 GAATTTTGTGATGAATAGTGACTCCATGCCCTTGAGAGTTGTCAGATGTTGGAGG 180	DB	389 CGTCTCTCCCTCCCTGCGGACCGGCTCGGCCCTCCCTGGGACGCGCCCTCCCTCCC 448
QY	70 GGCTCAGCAATGGGTGACACCGTCACGGTCCCTCCCTCCAGCGTCCCTCCGCCCTCCGCG 129	QY	89 ACCGTACGTGCCCTCCCTCCAGCGTCTCCGCCCTCCGCCCTCCGCCCTCCCTCCGCACTGT 148
DB	181 GGCTGCAGATCAGTGACACCTCTAAGTCCCTCTCTCTAGCCCTAGCCAGCCCTCTAC 240	DB	449 AGCGGACCGCTCTCTCTCCCTGGGACCTCGCTCTCTCTCTCTCTCTCTCTCTCTCT 508
QY	130 CCCCCTCTTGGCACTGTACATTCATATCATTTT 165	QY	149 AC 150
DB	241 CCCACCCCTCTAGGCTGAAGGGGTGATCCATTTT 276	DB	509 GC 510
RESULT 12			
ADC87092		RESULT 13	
ID	ADC87092 standard; DNA; 1141 BP.	ADA53147	
XX	AC	ID	ADA53147 standard; cDNA; 2179 BP.
XX	ADC87092;	AC	ADA53147;
DT	01-JAN-2004 (first entry)	XX	AC
XX	Human GPCR gene SEQ ID NO:1545.	DT	20-NOV-2003 (first entry)
DE	ds; gene; human; GPCR;	DE	Human coding sequence, SEQ ID 715.
KW	guanosine triphosphate-binding protein coupled receptor; gene therapy.	XX	Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KW		KW	Gene Therapy; human; secretory protein; membrane proteins; cancer;
OS	Homo sapiens.	KW	Inflammatory disease; osteoporosis; neurological disease; gene; ss.
XX	EP1270724-A2.	XX	Homo sapiens.
XX	02-JAN-2003.	XX	EP1293569-A2.
XX	18-JUN-2002; 2002EP-00013517.	XX	19-MAR-2003.
XX	18-JUN-2001; 2001JP-00246789.	XX	21-MAR-2002; 2002EP-00006586.
PA	(NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.	XX	14-SEP-2001; 2001JP-00328381.
PA	(ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.	PR	24-JAN-2002; 2002US-0350435P.
XX	Suwa M, Asai K, Akiyama Y, Aburatani H;	XX	(HELI-) HELIX RES INST.
XX	WPI; 2003-315783/31.	PA	(REAS-) RES ASSOC BIOTECHNOLOGY.
DR	P-PSDB; ADC87093.	XX	Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
XX	New polynucleotide, useful for preparing a composition for treating a	PI	Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PT	patient in need of increased or suppressed activity or expression of the	PI	Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
PT	guanosine triphosphate-binding protein coupled receptor.	XX	WPI; 2003-395539/38.
XX	Claim 1; SEQ ID NO 1545; 28pp; English.	DR	P-PSDB; ADA54786.
XX	The invention relates to a novel polynucleotide encoding a guanosine	XX	New polynucleotides encoding full-length polypeptides, e.g. secretory
CC	triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of	XX	and/or membrane proteins, useful for developing medicines for diseases in
CC	the invention may have a use in gene therapy. The polynucleotide and	PT	which the gene is involved, or as target molecules for gene therapy.
CC	polypeptide are useful for preparing a composition for treating a patient	XX	Claim 1; SEQ ID NO 715; 205pp; English.
CC	in need of increased or suppressed activity or expression of the	XX	The present invention relates to novel human secretory or membrane
CC	guanosine triphosphate-binding protein coupled receptor. The	CC	proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC	polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the	CC	ADA54071). The coding sequences are useful in the gene therapy of
		CC	diseases caused by abnormalities of the proteins, e.g. cancer,
		CC	inflammatory diseases, osteoporosis or neurological disease.
		XX	Sequence 2179 BP; 341 A; 737 C; 661 G; 440 T; 0 U; 0 Other;
		SQ	Query Match
			14.6%; Score 35.6; DB 7; Length 2179;
			Best Local Similarity 55.7%; Pred. No. 1.4;
			Matches 68; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

Qy	29	CGGTGATCTCATCTGCGGAGCGCGAAACTTTTGCAGGAGGCTACGCAATTGGCTGAC	88
Db	1163	CGTCTCCCTCCCTGCGGACCGCGCTCGGACCTCCCTGGGACCGCGCCCCCTCCCTCC	1222
Qy	89	ACCGTCACGTGCGCCCTCTCTCAGCGTCTCGGCCCTCCGCCGCCCTCTTGGCACTGT	148
Db	1223	AGCGGACCGCGTCTCTCCCTCCCTGGGACCTCGCTCCTCCCTCCCTCCCTGGCACTGT	1282
Qy	149	AC 150	
Db	1283	GC 1284	

RESULT 14	
ADB62471	
ID	ADB62471 standard; cDNA; 2818 bp.
XX	
AC	ADB62471;
XX	
DT	04-DEC-2003 (first entry)
XX	
DE	Human cDNA encoding clone FEBRA20078800.
XX	
KW	Human; ss; gene; pharmaceutical; diagnostic; gene therapy;
KW	tissue regeneration; cell regeneration; membrane protein;
KW	signal transduction-related protein; transcription-related protein;
KW	osteoporosis; neurological disease; cancer; tumour.
XX	
OS	Homo sapiens.

PH	Key	Location/Qualifiers
FT	CDS	24..644
FT		/*tag= a
FT		/product= "Clone FERRA20078800 protein"
XX		
PN	EP1308459-A2.	

PD	07-MAY-2003.
XX	
PF	28-MAR-2002; 2002EP-00007401.
XX	
PR	05-NOV-2001; 2001JP-00379298.
PR	25-JAN-2002; 2002US-00350978.
XX	
PA	(HELI-) HELIX RES INST.
PA	(REAS-) RES ASSOC BIOTECHNOLOGY.

PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;  
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;  
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;  
XX  
XX WPI; 2003-450961/43.  
DR  
DR P-PSDB; ADB64441.

New polynucleotides and polypeptides, useful for developing a diagnostic marker or medicines for regulation of their expression and activity, or as targets of gene therapy.

PS Claim 1; Page: 222pp: English.

The invention discloses a polynucleotide comprising a sequence selected from 1970 fully defined nucleotide sequences which encode novel polypeptides. Also claimed is a polypeptide encoded by the polynucleotide or its partial peptide, an antibody binding to the polypeptide or peptide of the polynucleotide, immunologically assaying the polypeptide or peptide of the polynucleotide by contacting the polypeptide or peptide with the antibody of the encoded protein, and observing the binding between the two, a transformant carrying the polynucleotide in an expressible manner and an antisense polynucleotide. The oligonucleotide is useful as a primer for synthesising the polynucleotide, or as a probe for detecting the polynucleotide. The polynucleotides and encoded proteins are useful as pharmaceutical agents and many disease-related

CC genes may be included in them, for developing a diagnostic marker or  
CC medicines for regulation of their expression and activity, or as targets  
CC of gene therapy. The genes are involved in tissue and/or cell  
CC regeneration. Membrane proteins, signal transduction-related proteins,  
CC transcription-related proteins, disease-related proteins and genes  
CC encoding them can be used as indicators for diseases (e.g. osteoporosis,  
CC neurological diseases, cancer, tumours). The cDNA may be used to regulate  
CC the activity or expression of the encoded protein to treat diseases. The  
CC sequence presented is a cDNA of the invention. Note: Some of the sequence  
CC data for this patent is not represented in the printed specification, but  
CC is based on sequence information supplied by the European Patent Office.  
XX  
SQ Sequence 2818 BP; 449 A; 956 C; 873 G; 540 T; 0 U; 0 Other:

Query Match	14.6%;	Score 35.6;	DB 9;	Length 2818;
Best Local Similarity	55.7%;	Pred. No. 1.6;		
Matches 68;	Conservative 0;	Mismatches 54;	Indels 0;	Gaps 0

  

QY	29	CGGTGATCCATCACTGCGGAAGCCGCAAACTTTTGACGAGGCTCAGCCATTTGGCTGCAC	88
Db	1798	CGCTCTCTCTCCCTGCGGACCGCGTTCGCGCCCTCCTTGGGAGCAGCGCCCTCCCTCCC	185
QY	89	ACCGTCACGTGCGCCCTCCTCAGCGTCTTCGCGCCCTCCCGCCCCCTCTTGGCAGCTGT	148
Db	1858	AGCGGACCGCGTCTCTCCCTCCCTGGGAGCTTCGCTCTTCTCCCTCCCTCGGACTGT	191
QY	149	AC 150	
Db	1918	GC 1919	

RESULT 15  
ADB62760  
ID ADB62760 standard; cDNA; 3089 BP.  
XX  
XX  
AC ADB62760;  
XX  
XX  
DT 04-DEC-2003 (first entry)  
XX  
XX  
DE Human cDNA encoding clone NT2NE20117580.

Human; ss; gene; pharmaceutical; diagnostic; gene therapy;  
tissue regeneration; cell regeneration; membrane protein;  
signal transduction-related protein; transcription-related protein;  
osteoporosis; neurological disease; cancer; tumour.  
XX  
OS Homo sapiens.

Accession	Key	Location/Qualifiers
FT	CDS	271..912
FT		/tag= a
FT		/product= "Clone NT2NE20117580 protein"
XX		
PN		EP1308459-A2.

XA	
PD	07-MAY-2003.
XX	
XX	28-MAR-2002; 2002EP-00007401.
XX	
XX	05-NOV-2001; 2001JP-00379298.
XX	25-JAN-2002; 2002US-00350978.
XX	
PA	(HELI-) HELIX RES INST.
PPA	(REAS-) RES ASSOC BIOTECHNOLOGY.
PPA	
XX	Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PPI	Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Tamechika I;
PPI	Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX	
XX	WPI; 2003-450961/43.
DR	
DR	P-PADB; ADB64730.
DR	
XX	
XX	New polynucleotides and polypeptides. useful for developing a diagnostic

Query Match	14.6%;	Score 35.6;	DB 9;	Length 3089;
Best Local Similarity	55.7%;	Pred. No. 1.6;		
Matches 68;	Conservative 0;	Mismatches 54;	Indels 0;	Gaps 0;
QY	29	CGGTGATCCATCACTGCGGAAGCGCGCAAACTTTTCAGGAGGCTCAGCCATTGGCTGAC	88	
Db	2066	CGCTCTCTCCCTCCCTGCGGACCGGCGCTCGGCGCTCCCTGGGACCGCGCCCTCCCTCC	2125	
QY	89	ACCGTCACGTGCCCCCTCTCTCAGCGTCTCTCGGCCCTCCCGCCCCCTCTTTGGCGACTGT	148	
Db	2126	AGCGGACCGGCGCTCCCTCCCTCGGGACCTCGTCTCTCCCTCCCTCGCGGACTGT	2185	
QY	149	AC 150		
Db	2186	GC 2187		

Search completed: September 18, 2004, 11:29:57  
Job time : 265.692 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 2004, 11:14:16 ; Search time 61.1263 Seconds  
(without alignments)  
2215.217 Million cell updates/sec

Title: US-10-614-282-2

Perfect score: 244

Sequence: 1 agcgccggggaatttaaatgc.....accaggtcaactgaagacag 244

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA.\*

- 1: /cgn2\_6/ptodata/2/ina/5A COMB.seq.\*
- 2: /cgn2\_6/ptodata/2/ina/5B COMB.seq.\*
- 3: /cgn2\_6/ptodata/2/ina/6A COMB.seq.\*
- 4: /cgn2\_6/ptodata/2/ina/6B COMB.seq.\*
- 5: /cgn2\_6/ptodata/2/ina/PCTUS COMB.seq.\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	15.2	4403765	3	US-09-103-840A-2
2	37	15.2	4411529	3	US-09-103-840A-1
3	35.4	14.5	7218	1	US-08-232-463-14
4	33.6	13.8	1778	4	US-08-934-386-4
5	32.4	13.3	315	4	US-09-712-529-6
6	32.4	13.3	12001	1	US-08-458-568A-11
7	32.4	13.3	34094	4	US-09-292-034-1
8	32	13.1	4190	3	US-08-924-345-1
9	31	12.7	4895	3	US-09-053-866-1
10	31	12.7	4895	4	US-09-479-130-1
11	31	12.7	4895	4	US-09-472-130A-1
12	30.8	12.6	1664976	4	US-08-916-421B-1
13	30.4	12.5	4403765	3	US-09-103-840A-2
14	30.4	12.5	4411529	3	US-09-103-840A-1
15	30	12.3	832	4	US-09-621-976-2813
16	29.8	12.2	12565	3	US-09-345-217-3
17	29.6	12.1	3226	3	US-08-870-126-10
18	29.6	12.1	3226	4	US-09-445-247-10
19	29.2	12.0	2451	4	US-09-252-991A-12219
20	29	11.9	432	4	US-09-252-991A-6795
21	29	11.9	556	3	US-09-058-489-25
22	29	11.9	816	4	US-09-252-991A-14049
23	29	11.9	933	4	US-09-252-991A-14130
24	29	11.9	1230	4	US-09-252-991A-6843
25	29	11.9	1278	4	US-09-252-991A-14006
26	29	11.9	1488	4	US-09-252-991A-6748
27	29	11.9	1773	4	US-09-252-991A-6876

28	28.8	11.8	448	4	US-09-621-976-1871	Sequence 1871, Ap
29	28.8	11.8	2861	4	US-09-016-434-1103	Sequence 1103, Ap
30	28.8	11.8	3510	3	US-09-056-105-16	Sequence 16, Appl
31	28.8	11.8	31571	1	US-08-323-443B-1	Sequence 1, Appli
32	28.8	11.8	33529	3	US-09-144-085-3	Sequence 3, Appli
33	28.6	11.7	11220	3	US-09-105-537-32	Sequence 32, Appl
34	28.6	11.7	36778	3	US-09-105-537-5	Sequence 5, Appli
35	28.6	11.7	38506	3	US-09-320-878-19	Sequence 19, Appl
36	28.6	11.7	38506	4	US-09-141-908-1	Sequence 1, Appli
37	28.6	11.7	38506	4	US-09-657-440-19	Sequence 19, Appli
38	28.6	11.7	53526	3	US-08-658-136-2	Sequence 2, Appli
39	28.6	11.7	53577	3	US-08-658-136-1	Sequence 1, Appli
40	28.4	11.6	753	4	US-09-489-039A-1736	Sequence 1736, Ap
41	28.4	11.6	1227	4	US-09-620-312D-1027	Sequence 1027, Ap
42	28.4	11.6	1380	4	US-09-620-312D-1026	Sequence 1026, Ap
43	28.4	11.6	2301	1	US-08-306-691B-23	Sequence 23, Appl
44	28.4	11.6	2301	4	US-09-167-206-3	Sequence 3, Appli
45	28.4	11.6	2301	5	PCT-US93-06251-78	Sequence 78, Appl

#### ALIGNMENTS

##### RESULT 1

US-09-103-840A-2  
; Sequence 2, Application US/09103840A  
; Patent No. 6294328

; GENERAL INFORMATION:

; APPLICANT: FLEISCHMAN, Robert D.

; APPLICANT: WHITE, Owen R.

; APPLICANT: FRASER, Claire M.

; APPLICANT: VENTER, John C.

; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

; FILE REFERENCE: 24366-20007.00

; CURRENT APPLICATION NUMBER: US/09/103,840A

; CURRENT FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 4403765

; TYPE: DNA

; ORGANISM: Mycobacterium tuberculosis

; FEATURE:

; OTHER INFORMATION: CDC 1551

; OTHER INFORMATION: "n" bases at various positions throughout the sequence

; OTHER INFORMATION: represent a, t, c or g

US-09-103-840A-2

Query Match

Best Local Similarity 15.2%; Score 37; DB 3; Length 4403765;

Matches 82; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 45 CGGAAGCGGCAAACTTTTCAGAGAGGCTTCAGCATGTCGACCGTACCGTCCCT 104

Db 2520637 CGTGGCGGCTCACAATGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2520696

QY 105 CTTCAGCGGCTTCCTCCGCGGCTTCCTCCGCGGCTTCCTCCGCGGCTTCCTCCG 164

Db 2520697 CCACGAGGACTCGGCAATCCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2520756

QY 165 TCTTCTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 201

Db 2520757 CGCCGATCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2520793

RESULT 2

US-09-103-840A-1

; Sequence 1, Application US/09103840A

; Patent No. 6294328

; GENERAL INFORMATION:

; APPLICANT: FLEISCHMAN, Robert D.

; APPLICANT: WHITE, Owen R.

APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
STRANDEDNESS: TUBERCULOSIS  
FILE REFERENCE: 24366-20007.00  
CURRENT APPLICATION NUMBER: US/09/103,840A  
CURRENT FILING DATE: 1998-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 4411529  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
OTHER INFORMATION: H3/RV  
US-09-103-840A-1

Query Match 15.2%; Score 37; DB 3; Length 4411529;  
Best Local Similarity 52.2%; Pred. No. 0.6; Indels 0; Gaps 0;  
Matches 82; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 45 CGGAAGCCGCGAACTTTTCAGGAGGCTGACCCATTGGCTGACACCGCTACGTCGCCCT 104  
DB 2523409 CGGTGGCGGCTCACATGTGGCGGCGGCGAGCGACGTGGCCGCTCGGCCGTACCGCG 2523468  
QY 105 CTTCCAGGCTCTCGCCCTCCGCCCTTTCGGCACTGTACATTCATATTTT 164  
DB 2523469 CCACGAGCACTCGGCGCAATCCGGCGCCGATCCGGGGCGCCGCGAGGTTTCGCCGGTG 2523528  
QY 165 TCTTCTCCGCGCCCTGAGGAGGAGTGAGAAAGTTGGC 201  
DB 2523529 CCGCGATCAGCGGCGAGTTCGAGTGGCGCACTTCGC 2523565

RESULT 3  
US-08-232-463-14  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZgpt-F1s  
US-08-232-463-14

Query Match 14.5%; Score 35.4; DB 1; Length 7218;  
Best Local Similarity 6.7%; Pred. No. 0.12;  
Matches 9; Conservative 85; Mismatches 41; Indels 0; Gaps 0;

QY 44 CGGGAAGCCGCGAACTTTTCAGGAGGCTGACCCATTGGCTGACACCGTACGTCGCCCT 103  
DB 1052 GAGGAGCTTCGATTT 1111  
QY 104 TCCTCAGGCTCTCGCCCTCCGCCCTTTCGGCACTGTACATTCATATTCATTT 163  
DB 1112 YY 1171  
QY 164 TCTTCTCCGCGCCCT 178  
DB 1172 YYYYYYYYYYYYYY 1186

RESULT 4  
US-08-934-386-4  
Sequence 4, Application US/08934386  
Patent No. 6306636  
GENERAL INFORMATION:  
APPLICANT: Haselkorn, Robert  
APPLICANT: Gornicki, Piotr  
TITLE OF INVENTION: Methods for Detecting Nucleic Acid  
TITLE OF INVENTION: Segments Encoding Acetyl-CoA Carboxylase  
NUMBER OF SEQUENCES: 68  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 433  
CITY: Houston  
STATE: Texas  
COUNTRY: US  
ZIP: 77210-4433  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/934,386  
FILING DATE: 19-SEP-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Kitchell, Barbara S.  
REGISTRATION NUMBER: 33,928  
REFERENCE/DOCKET NUMBER: ARSB:521  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 418-3000  
TELEFAX: (713) 789-2679  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1778 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-934-386-4

Query Match 13.8%; Score 33.6; DB 4; Length 1778;  
Best Local Similarity 55.0%; Pred. No. 0.25; Indels 0; Gaps 0;  
Matches 66; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 42 CTGCGGAAGCCGCGAACTTTTCAGGAGGCTGACCATTTGGCTGACCGTCACGTCACGTGCC 101

Db 349 CCGCCCAAGCAAGCAGCGCCATGAGGACCACTACCGTCTGTCCTCCGTAGACCA 408  
QY 102 CCTCTCAGGCTCTCGGCTCCGCGCCCTCTTGGCAGCTGTACATTCATATCAT 161  
Db 409 CACCCCACTGCGCCCACTCTCTCTCCCACTCTCTCATGAATGACGCGATTCAT 468

## RESULT 5

US-09-712-529-6/c

; Sequence 6, Application US/09712529  
; Patent No. 6485938  
; GENERAL INFORMATION:  
; APPLICANT: Sheppard, Paul O.  
; APPLICANT: Bishop, Paul D.  
; APPLICANT: Whitmore, Theodore E.  
; APPLICANT: Thompson, Penny P.  
; TITLE OF INVENTION: Human Zven Proteins  
; FILE REFERENCE: 99-81  
; CURRENT APPLICATION NUMBER: US/09/712,529  
; CURRENT FILING DATE: 2000-11-14  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 6  
; LENGTH: 315  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: This degenerate sequence encodes the amino acid  
; NAME/KEY: misc feature  
; LOCATION: (1)...(315)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-712-529-6

Query Match 13.3%; Score 32.4; DB 4; Length 315;  
Best Local Similarity 33.6%; Pred. No. 0.3;  
Matches 44; Conservative 22; Mismatches 65; Indels 0; Gaps 0;  
QY 24 TAAACGGTGATCCACTCAGGAGCGGCAACTTTTGCAGGAGGCTCAGCATGG 83  
Db 182 TGCAYTCYTCNCCYTCNKNCCNARNGNGTRCACATNCKNARNCNCCNARNCANRN 123  
QY 84 CTGACACGTCACGTCGCTCCCTCTCCAGCGTCCTCCGCGCTCCGCGCTCTCTGCGC 143  
Db 122 SWDATNGRCARCANGTNCNGCNCRCRCAYTGNACRTCNCKYTCRCANGCNCNGTDATN 63  
QY 144 ACTGTATC 154  
Db 62 ACNGRCATC 52

## RESULT 6

US-08-458-568A-11/c  
; Sequence 11, Application US/08458568A  
; Patent No. 5821339  
; GENERAL INFORMATION:  
; APPLICANT: Schaffer, Priscilla A.  
; APPLICANT: Yeh, Lily  
; TITLE OF INVENTION: Compositions and Methods for Treatment of Herpesvirus  
; TITLE OF INVENTION: Infections  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5821339ris  
; STREET: One Liberty Place, 46th floor  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19103  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WordPerfect 5.1

; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/458,568A  
; FILING DATE: 02-JUNE-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/065,146  
; FILING DATE: 05-MAY-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Leary Ph.D., Kathryn R.  
; REGISTRATION NUMBER: 36,317  
; REFERENCE/DOCKET NUMBER: DFCI-0029  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (215) 568-3100  
; TELEFAX: (215) 568-3439  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Herpes simplex virus  
; STRAIN: Herpes Simplex Virus Type 1  
US-08-458-568A-11

Query Match 13.3%; Score 32.4; DB 1; Length 12001;  
Best Local Similarity 64.9%; Pred. No. 1.5;  
Matches 48; Conservative 0; Mismatches 26; Indels 0; Gaps 0;  
QY 72 CTCAGCCATTTGGCTGACACCGTCACGTGCCCCCTCTCTCCAGGCTCTCCGCGCTCCCGCCC 131  
Db 7690 CTCGCGCCCGCGTCCCTCTCTCCGCGCCCGCGTCCCTCTCTCCGCGCGCGCTCC 7631  
QY 132 CCCCTCTTGGGCAC 145  
Db 7630 CCCCTCTCTCGGCC 7617

## RESULT 7

US-09-292-034-1/c  
; Sequence 1, Application US/09292034  
; Patent No. 6492343  
; GENERAL INFORMATION:  
; APPLICANT: Reddy, P. Seshidhar  
; APPLICANT: Tikoo, Suresh  
; APPLICANT: Babiuk, Lorne  
; TITLE OF INVENTION: PORCINE ADENOVIRUS TYPE 3 GENOME  
; FILE REFERENCE: 293102002400  
; CURRENT APPLICATION NUMBER: US/09/292,034  
; CURRENT FILING DATE: 1999-04-14  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 34094  
; TYPE: DNA  
; ORGANISM: Porcine Adenovirus Type 3  
; FEATURE:  
US-09-292-034-1

Query Match 13.3%; Score 32.4; DB 4; Length 34094;  
Best Local Similarity 48.9%; Pred. No. 2.3;  
Matches 87; Conservative 0; Mismatches 91; Indels 0; Gaps 0;  
QY 65 CAGGAGGCTCAGCCATTTGGCTGACACCGTCACGTGCCCCCTCTCCAGGCTCTCCGCGCT 124  
Db 3204 CGGGGGGCGCGCAGCGCTGTGTCACTTTCATCTGCGCACCTGCGCGCCAGGGGTGATTG 3145  
QY 125 CCGCGCCCCCTCTTGGGACTGTATTCATCATATCATTTTCTTCTCCGCCCCCATGGAG 184

Db 3144 TCCCGCAGCAGCGGCTACTGCTGAGAGGTATCCGTCTCCATCTCGTCGCTCCGTA 3085  
QY 185 GAAGTGAAGATTGGCGACAGTCACGCCGGGCTTCGAGGACCAAGTCAGTCAGTGAC 242  
Db 3084 CACTGCAGCAGCTTGTACATCTTCACAGGTCTGATCGAAGCAGCAGGTCACTCAGAC 3027

## RESULT 8

US-08-924-345-1  
; Sequence 1, Application US/08924345  
; Patent No. 6224878  
; GENERAL INFORMATION:  
; APPLICANT: LEUNG-TACK Patricia  
; APPLICANT: LEGATELOIS Isabelle, Christine, Marie-Andree  
; APPLICANT: AUDONNET Jean-Christophe, Francis  
; APPLICANT: RIVIERE Michel, Emile, Albert  
; TITLE OF INVENTION: Mutants and vaccines of the Infectious  
; TITLE OF INVENTION: Bovine Rhinotracheitis virus  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: LARSON AND TAYLOR  
; STREET: 727 SOUTH TWENTY-THIRD STREET  
; CITY: ARLINGTON  
; STATE: VIRGINIA  
; COUNTRY: USA  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/924,345  
; FILING DATE: 04-SEP-1997  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/199,172  
; FILING DATE: 09-AUG-1994  
; APPLICATION NUMBER: FR 92 07930  
; FILING DATE: 26-JUN-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: SARRO, THOMAS P.  
; REGISTRATION NUMBER: 19396  
; REFERENCE/DOCKET NUMBER: XI  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 920-7200  
; TELEFAX: (703) 892-8428  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4190 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Type 1 Bovine Herpesvirus  
; STRAIN: ST  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 172..1311  
; OTHER INFORMATION: /function= "envelope glycoprotein"  
; OTHER INFORMATION: /product= "glycoprotein g1"  
; OTHER INFORMATION: /standard\_name= "BHV-1 g1"  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1594..3318  
; OTHER INFORMATION: /function= "envelop glycoprotein"  
; OTHER INFORMATION: /product= "glycoprotein gE"  
; OTHER INFORMATION: /standard\_name= "BHV-1 gE"  
; FEATURE:  
; NAME/KEY: CDS

; LOCATION: 3388..3861  
; OTHER INFORMATION: /product= "protein US9"  
; OTHER INFORMATION: /standard\_name= "BHV US9"  
US-08-924-345-1  
Query Match 13.1%; Score 32; DB 3; Length 4190;  
Best Local Similarity 52.2%; Pred. No. 1.3;  
Matches 71; Conservative 0; Mismatches 65; Indels 0; Gaps 0;  
QY 42 CTGCGGAAGCCGGCAAACTTTTGAGAGGCTCAGCCATTGGCTCAGACCGTCAGTCAGTCGCC 101  
Db 3818 CAGCGGACGGCGCGCTCTCTGGCTGTGGAGGTAGTGCCCGCGCCCTCAGGCGGGCGCCGAC 3877  
QY 102 CCTCTCCAGGTCTCTCCGCCCTCCGCCGCCCTCTTTGGCGACTGTATCATTCATCAT 161  
Db 3878 TGTCCCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 3937  
QY 162 TTTTCTCTCCGGGCC 177  
Db 3938 TTTTTCGACGCGCC 3953  
RESULT 9  
US-09-053-866-1  
; Sequence 1, Application US/09053866  
; Patent No. 6111075  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Wenfeng  
; APPLICANT: Presnell, Scott R.  
; APPLICANT: Yee, David P.  
; APPLICANT: Foster, Donald C.  
; TITLE OF INVENTION: PROTEASE-ACTIVATED RECEPTOR  
; TITLE OF INVENTION: PAR4 (ZCHEMR2)  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: ZymoGenetics, Inc.  
; STREET: 1201 Eastlake Avenue East  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98102  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/053,866  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Leith, Debra K  
; REGISTRATION NUMBER: 32,619  
; REFERENCE/DOCKET NUMBER: 98-10  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 206-442-6674  
; TELEFAX: 206-442-6678  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4895 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: Coding Sequence  
; LOCATION: 176...1330  
; OTHER INFORMATION:  
; US-09-053-866-1



```

Query Match      12.7%; Score 31; DB 3; Length 4895;
Best Local Similarity 56.3%; Pred. No. 2.9;
Matches 58; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY      98  TGCCCTCCCTCCAGCGTCTCCGCCCTCCGCCCTCCGCCCTCTTGGCACTGTACATTGATA 157
      Db      3108  TCCCTGACTGCTGGGACCCCTCGCCCTTCCTCTCTCACCTGCAGGCTGATCCTCTTT 3167

QY      158  TCAATTTTCTCTCCGGCCCCCATGGAGGAGTGGAGAAAGTTGG 200
      Db      3168  TCACTTTCTGTCAATGTCAACAGGAGATAAGGTGGACAAATGGG 3210

RESULT 10
US-09-479-130-1
; Sequence 1, Application US/09479130
; Patent No. 6436400
; GENERAL INFORMATION:
; APPLICANT: Xu, Wenfeng
; APPLICANT: Presnell, Scott R.
; APPLICANT: Yee, David P.
; APPLICANT: Foster, Donald C.
; TITLE OF INVENTION: PROTEASE-ACTIVATED RECEPTOR
; TITLE OF INVENTION: PAR4 (ZCHEMR2)
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ For Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/479,130
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Leith, Debra K
; REGISTRATION NUMBER: 32,619
; REFERENCE/DOCKET NUMBER: 98-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6674
; TELEFAX: 206-442-6678
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4895 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 176...1330
; OTHER INFORMATION:
; US-09-479-130-1

Query Match      12.7%; Score 31; DB 4; Length 4895;
Best Local Similarity 56.3%; Pred. No. 2.9;
Matches 58; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY      98  TGCCCTCCCTCCAGCGTCTCCGCCCTCCGCCCTCCGCCCTCTTGGCACTGTACATTGATA 157
      Db      3108  TCCCTGACTGCTGGGACCCCTCGCCCTTCCTCTCTCACCTGCAGGCTGATCCTCTTT 3167

QY      158  TCAATTTTCTCTCCGGCCCCCATGGAGGAGTGGAGAAAGTTGG 200

```

```

; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98373)..(98373)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (10398)..(10398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g

```

NAME/KEY: misc_feature	
LOCATION: (622708)..(622708)	
OTHER INFORMATION: n equals a,	t, c, or g
NAME/KEY: misc_feature	
LOCATION: (657081)..(657081)	
OTHER INFORMATION: n equals a,	t, c, or g
NAME/KEY: misc_feature	
LOCATION: (657203)..(657203)	
OTHER INFORMATION: n equals a,	t, c, or g
NAME/KEY: misc_feature	
LOCATION: (674435)..(674435)	
OTHER INFORMATION: n equals a,	t, c, or g
NAME/KEY: misc_feature	
LOCATION: (682442)..(682442)	
OTHER INFORMATION: n equals a,	t, c, or g
NAME/KEY: misc_feature	
LOCATION: (713652)..(713652)	
OTHER INFORMATION: n equals a,	t, c, or g
NAME/KEY: misc_feature	
LOCATION: (741684)..(741684)	
OTHER INFORMATION: n equals a,	t, c, or g
NAME/KEY: misc_feature	
LOCATION: (779455)..(779455)	
OTHER INFORMATION: n equals a,	t, c, or g
NAME/KEY: misc_feature	
LOCATION: (779676)..(779676)	
OTHER INFORMATION: n equals a,	t, c, or g
NAME/KEY: misc_feature	
LOCATION: (855539)..(855539)	
OTHER INFORMATION: n equals a,	t, c, or g
NAME/KEY: misc_feature	
LOCATION: (871619)..(871619)	
OTHER INFORMATION: n equals a,	t, c, or g
NAME/KEY: misc_feature	
LOCATION: (1084830)..(1084830)	
OTHER INFORMATION: n equals a,	t, c, or g
NAME/KEY: misc_feature	
LOCATION: (1096846)..(1096846)	
OTHER INFORMATION: n equals a,	t, c, or g
NAME/KEY: misc_feature	
LOCATION: (1119881)..(1119881)	
OTHER INFORMATION: n equals a,	t, c, or g
NAME/KEY: misc_feature	
LOCATION: (1130881)..(1130881)	
OTHER INFORMATION: n equals a,	t, c, or g
NAME/KEY: misc_feature	
LOCATION: (1310988)..(1310988)	
OTHER INFORMATION: n equals a,	t, c, or g
NAME/KEY: misc_feature	
LOCATION: (1313224)..(1313224)	
OTHER INFORMATION: n equals a,	t, c, or g
NAME/KEY: misc_feature	
LOCATION: (1349473)..(1349473)	
OTHER INFORMATION: n equals a,	t, c, or g
NAME/KEY: misc_feature	
LOCATION: (1349491)..(1349491)	
OTHER INFORMATION: n equals a,	t, c, or g
NAME/KEY: misc_feature	
LOCATION: (1470091)..(1470091)	
OTHER INFORMATION: n equals a,	t, c, or g
NAME/KEY: misc_feature	
LOCATION: (1569020)..(1569020)	
OTHER INFORMATION: n equals a,	t, c, or g
NAME/KEY: misc_feature	
LOCATION: (1602912)..(1602912)	
OTHER INFORMATION: n equals a,	t, c, or g
NAME/KEY: misc_feature	
LOCATION: (1603734)..(1603734)	
OTHER INFORMATION: n equals a,	t, c, or g
NAME/KEY: misc_feature	
LOCATION: (1637998)..(1637998)	
OTHER INFORMATION: n equals a,	t, c, or g
NAME/KEY: misc_feature	

```

; LOCATION: (1664854)..(1664855)
; OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

Query Match
Best Local Similarity 12.6%; Score 30.8; DB 4; Length 1664976;
Matches 50; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 96 CQTGCCCCCTCTCCAGAGCTCTCCGCGCTCCGCGCCCTCTTGGCACTGTATCA 155
Db 919539 CCTCCTCGCCCCAGATCTCTTCCGTCGCGACCCCGCGAGCGCGCTCTGCC 919598

QY 156 TATCATTTTCTCTCCGCGCC 177
Db 919599 TACCGTGTCTCTCCGCGCC 919620

RESULT 13
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103.840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match
Best Local Similarity 12.5%; Score 30.4; DB 3; Length 4403765;
Matches 73; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 88 CACGTCACGTGCCCCCTCTCCAGCTCTCTCCGCGCTCCGCGCCCTCTTGGCACTG 147
Db 2062823 CACCGGAATGTGCGGTCTCTCGACAAGACGACGCAATGCGCTCCGCAATGCGCCACCA 2062764

QY 148 TACATTCATATCATTTTCTCTCCGCGCCCATGAGAGTGAAGAGTTGGCAGATC 207
Db 2062763 GACGTGGTACCTGCCATGATGACCGACGCTGGCGCTCGAAAGACGCGCGCCGCG 2062704

QY 208 ACGCCGGGCTTCGAGGACAGGT 231
Db 2062703 GTCCAGGGCGTCCGCGCTGGGT 2062680

RESULT 14
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103.840A

```

```

; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37RV
US-09-103-840A-1

Query Match
Best Local Similarity 12.5%; Score 30.4; DB 3; Length 4411529;
Matches 73; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 88 CACGTCACGTGCCCCCTCTCCAGCTCTCTCCGCGCTCCGCGCCCTCTTGGCACTG 147
Db 2065516 CACCGGAATGTGCGGTCTCTCGACAAGACGACGCAATGCGCTCCGCAATGCGCCACCA 2065457

QY 148 TACATTCATATCATTTTCTCTCCGCGCCCATGAGAGTGAAGAGTTGGCAGATC 207
Db 2065456 GACGTGGTACCTGCCATGATGACCGGACGCTGGCGCTCGAAAGACGCGCGCGCCG 2065397

QY 208 ACGCCGGGCTTCGAGGACAGGT 231
Db 2065396 GTCCAGGGCGTCCGCGCTGGGT 2065373

```

```

RESULT 15
US-09-621-976-2813/c
; Sequence 2813, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621.976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 2813
; LENGTH: 832
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 235..399
US-09-621-976-2813

Query Match
Best Local Similarity 12.3%; Score 30; DB 4; Length 832;
Matches 18; Conservative 95; Mismatches 75; Indels 0; Gaps 0;

QY 11 ATTATAATGCCATAAAACGTCATCATCATCTCGGAAGCGGCAAACTTTGACGAG 70
Db 357 WKTWYWTYTTWTRMMMKKARRWYWMKSTYACASRYRKYTWGWWYMKMMSTRY 298

QY 71 GCTCAGCATTTGGCTGACACCGTCAGTCCCTCTCCAGCGTCTCCGCCCTCCGCC 130
Db 297 CYMKCKCMYGRRCRWYTWARGMWSYANGKWSMSMCTRYKGYSTYWTWKCT 238

QY 131 CCCCCTCTTGGCACTGTATCATATCATTTTCTCTCCGCGCCCATGAGGAGTG 190
Db 237 CATWCYWKYKRWKSKTCSGSGRGYMTSYSTRSYMWSWYTWCMWWRWSTY 178

QY 191 AGAAAGTT 198
Db 177 WYMAWKK 170

```

Search completed: September 18, 2004, 13:19:20  
Job time : 83.1263 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 2004, 12:08:47 : Search time 269,259 Seconds  
(without alignments)  
4573.952 Million cell updates/sec

Title: US-10-614-282-2

Perfect score: 244

Sequence: 1 agcgcggggatttaaatgc.....accaggctcaactgacag 244

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3327077 seqs, 2523723180 residues

Total number of hits satisfying chosen parameters: 6654154

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/PCT NEW PUB.seq.\*
- 2: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW PUB.seq.\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW PUB.seq.\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW PUB.seq.\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW PUB.seq.\*
- 13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW PUB.seq.\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*
- 17: /cgn2\_6/ptodata/2/pubpna/US10\_NEW PUB.seq.\*
- 18: /cgn2\_6/ptodata/2/pubpna/US60\_NEW PUB.seq.\*
- 19: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	244	100.0	244	17	US-10-614-282-2
c 2	38	15.6	302250	9	US-09-962-832-154
c 3	37.2	15.2	288	13	US-10-424-599-116935
4	37.2	15.2	3050	16	US-10-104-047-262
5	37	15.2	63155	16	US-10-292-798-449
6	36.2	14.8	89829	17	US-10-322-281-359
7	36	14.8	374	9	US-09-867-701-4884
8	35.6	14.6	1141	15	US-10-017-161-1889
9	35.6	14.6	1141	16	US-10-292-798-1545
10	35.6	14.6	2179	16	US-10-094-749-715
11	35.6	14.6	2818	16	US-10-104-047-625
12	35.6	14.6	3089	16	US-10-104-047-914
13	35.6	14.6	3743	16	US-10-104-047-273
14	35.6	14.6	34875	17	US-10-775-169-316

c 15	35.4	14.5	900	13	US-10-424-599-141831	Sequence 141831, A
c 16	34.6	14.2	1616	17	US-10-437-963-78180	Sequence 78180, A
c 17	34.6	14.2	1616	17	US-10-437-963-78180	Sequence 78180, A
c 18	34.6	14.2	4739	9	US-09-954-531-586	Sequence 586, App
c 19	34.6	14.2	4739	13	US-10-193-651-22	Sequence 22, Appl
c 20	34.2	14.0	104514	13	US-10-087-192-1981	Sequence 1981, Ap
c 21	34	13.9	536	9	US-09-796-652-5305	Sequence 5305, Ap
c 22	34	13.9	536	15	US-10-040-862-5305	Sequence 5305, Ap
c 23	34	13.9	536	16	US-10-057-475B-5305	Sequence 5305, Ap
c 24	34	13.9	536	16	US-10-154-884B-5305	Sequence 5305, Ap
c 25	34	13.9	536	17	US-10-764-324-5305	Sequence 5305, Ap
c 26	33.4	13.7	1681	13	US-10-425-114-27309	Sequence 27309, A
c 27	33.4	13.7	2537	13	US-10-466-164-25	Sequence 25, Appl
c 28	33.4	13.7	16161	10	US-09-957-956-8	Sequence 8, Appl
c 29	33.2	13.6	731	13	US-10-027-632-153404	Sequence 153404, A
c 30	33.2	13.6	731	16	US-10-027-632-153404	Sequence 153404, A
c 31	32.8	13.4	875	17	US-10-767-701-12676	Sequence 12676, A
c 32	32.8	13.4	1004	17	US-10-437-963-26175	Sequence 26175, A
c 33	32.8	13.4	1800	16	US-10-369-493-44508	Sequence 44508, A
c 34	32.8	13.4	1950	17	US-10-437-963-42029	Sequence 42029, A
c 35	32.8	13.4	33137	16	US-10-292-798-913	Sequence 913, App
c 36	32.8	13.4	48841	10	US-09-844-653-32	Sequence 32, Appl
c 37	32.6	13.4	540	13	US-10-424-599-11898	Sequence 11898, A
c 38	32.4	13.3	315	15	US-10-212-355-6	Sequence 6, Appl
c 39	32.4	13.3	315	15	US-10-212-355-6	Sequence 6, Appl
c 40	32.4	13.3	315	17	US-10-680-755A-6	Sequence 6, Appl
c 41	32.4	13.3	315	17	US-10-680-800A-6	Sequence 6, Appl
c 42	32.4	13.3	492	13	US-10-424-599-59619	Sequence 59619, A
c 43	32.4	13.3	1959	16	US-10-260-238-1522	Sequence 1522, Ap
c 44	32.4	13.3	2097	16	US-10-369-493-28211	Sequence 28211, A
c 45	32.4	13.3	2115	16	US-10-369-493-30970	Sequence 30970, A

#### ALIGNMENTS

#### RESULT 1

US-10-614-282-2  
; Sequence 2, Application US/10614282  
; Publication No. US20040082034A1  
; GENERAL INFORMATION:  
; APPLICANT: LEE, DUNG-FANG  
; APPLICANT: JUANG, JYH-LYH  
; TITLE OF INVENTION: INTERNAL RIBOSOME ENTRY SITE OF THE LABIAL GENE FOR  
; FILE REFERENCE: 08842.0001-00000  
; CURRENT APPLICATION NUMBER: US/10/614,282  
; CURRENT FILING DATE: 2003-07-08  
; PRIOR APPLICATION NUMBER: 60/394,270  
; PRIOR FILING DATE: 2002-07-09  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 244  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-614-282-2

Query Match	100.0%;	Score 244;	DB 17;	Length 244;
Best Local Similarity	100.0%;	Pred. No. 5.5e-75;		
Matches 244;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	AGCGCGGGGATTAAATGCCACTAAAACGGTGATCCATCCTGCGAAGCGGCAAACT	60	
Db	1	AGCGCGGGGATTAAATGCCACTAAAACGGTGATCCATCCTGCGAAGCGGCAAACT	60	
QY	61	TTTGCAGGAGCTCAGCCATTGCTGACACCGTCAGTCCCTCTCTCAGCGTCTTCG	120	
Db	61	TTTGCAGGAGCTCAGCCATTGCTGACACCGTCAGTCCCTCTCTCAGCGTCTTCG	120	
QY	121	CCCTCCGCGCCCGCTTGGCACGTACATTCATATTTCTTCTCCGCGCCCAT	180	
Db	121	CCCTCCGCGCCCGCTTGGCACGTACATTCATATTTCTTCTCCGCGCCCAT	180	

```

RESULT 3
US-10-424-599-116935/c
/ Sequence 116935, Application US/10424599
/ Publication No. US20040031072A1
/ GENERAL INFORMATION:
/ APPLICANT: La Rosa Thomas J
/ APPLICANT: Kovalic David K
/ APPLICANT: Zhou Yihua
/ APPLICANT: Cao Yongwei
/ TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
/ TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
/ FILE REFERENCE: 38-21(53223)B
/ CURRENT APPLICATION NUMBER: US/10/424,599
/ CURRENT FILING DATE: 2003-04-28
/ NUMBER OF SEQ ID NOS: 285684
/ SEQ ID NO 116935
/ LENGTH: 288
/ TYPE: DNA
/ ORGANISM: Glycine max
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (1)..(288)
/ OTHER INFORMATION: unsure at all n locations:

```









; NUMBER OF SEQ ID NOS: 4096  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 273  
; LENGTH: 3743  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-104-047-273

Query Match 14.6%; Score 35.6; DB 16; Length 3743;  
Best Local Similarity 55.7%; Pred. No. 0.11;  
Matches 68; Conservative 0; Mismatches 54; Indels 0; Gaps 0;  
QY 29 CGGTGATCCATCACTCGGAGCGCGCAAACTTTTCAGAGGCTCAGGCATTGGCTGAC 88  
Db 2720 CGCTCTCTCTCTCTCGGACCGCGCTCGGCCCTCCCTGGGACCGCGCCCTCCCTCC 2779  
QY 89 ACCGTACAGTGGCCCTCTCTCCAGCGTCCTCGGCCCTCCCGCCCTCTTGGGCACTGT 148  
Db 2780 AGCGGACCGCGCTCTCTCCCTCCCTGGGACCTCGCTCTCTCCCTCCCTCGGACTGT 2839  
QY 149 AC 150  
Db 2840 GC 2841

## RESULT 14

US-10-775-169-316  
; Sequence 316, Application US/10775169  
; Publication No. US20040175743A1  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth  
; APPLICANT: Burczynski, Michael  
; APPLICANT: Twine, Natalie  
; APPLICANT: Dörner, Andrew  
; APPLICANT: Trepicchio, William  
; TITLE OF INVENTION: Method for Monitoring Drug Activities In Vivo  
; FILE REFERENCE: AM101080 (031896-013000)  
; CURRENT APPLICATION NUMBER: US/10/775,169  
; CURRENT FILING DATE: 2004-02-11  
; NUMBER OF SEQ ID NOS: 5278  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 316  
; LENGTH: 34875  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-775-169-316

Query Match 14.6%; Score 35.6; DB 17; Length 34875;  
Best Local Similarity 55.7%; Pred. No. 0.23;  
Matches 68; Conservative 0; Mismatches 54; Indels 0; Gaps 0;  
QY 29 CGGTGATCCATCACTCGGAGCGCGCAAACTTTTCAGAGGCTCAGGCATTGGCTGAC 88  
Db 18170 CGCTCTCTCTCTCTCGGACCGCGCTCGGCCCTCCCTGGGACCGCGCCCTCCCTCC 18229  
QY 89 ACCGTACAGTGGCCCTCTCTCCAGCGTCCTCGGCCCTCCCGCCCTCTTGGGCACTGT 148  
Db 18230 AGCGGACCGCGCTCTCTCTCCCTCGGAGCTCGCTCTCTCCCTCCCTCGGACTGT 18289  
QY 149 AC 150  
Db 18290 GC 18291

## RESULT 15

US-10-424-599-141831/c  
; Sequence 141831, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 141831  
; LENGTH: 900  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_99085C.1  
US-10-424-599-141831

Query Match 14.5%; Score 35.4; DB 13; Length 900;  
Best Local Similarity 57.8%; Pred. No. 0.084;  
Matches 63; Conservative 0; Mismatches 46; Indels 0; Gaps 0;  
QY 22 ACTAAAACGGTGATCCATCACTGCGGAGCGCGCAAACTTTTGCAGGAGGCTCAGCCATT 81  
Db 479 ACGAGACGCAATCCCAATTTGCATAGCAGGACCCCAATTTCTGAGCCTGAATCAGG 420  
QY 82 GGCTGACACCGTCAGTGCCCTCTCTCCAGCGTCTCTCGCCCTCCCGCC 130  
Db 419 GGCTGCCAACGCCACCGGGCGCACCGTGACCACCGCCCTTCTCTCC 371

Search completed: September 18, 2004, 15:14:59  
Job time : 271.259 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 2004, 10:33:16 : Search time 2061.12 Seconds  
(without alignments)  
3535.155 Million cell updates/sec

Title: US-10-614-282-2

Perfect score: 244

Sequence: 1 agccgggggttaaatgc.....accaggtcaactgaagacag 244

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_hic:\*

9: gb\_estl:\*

10: gb\_est2:\*

11: gb\_hic:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: em\_gss\_hum:\*

18: em\_gss\_inv:\*

19: em\_gss\_pln:\*

20: em\_gss\_vrt:\*

21: em\_gss\_fun:\*

22: em\_gss\_mam:\*

23: em\_gss\_mus:\*

24: em\_gss\_pro:\*

25: em\_gss\_rod:\*

26: em\_gss\_phg:\*

27: em\_gss\_vri:\*

28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	125.8	51.6	421	10	BE649209
2	124.2	50.9	343	10	AW121149
3	88	36.1	625	12	BM838150
4	80	32.8	2475	11	AK083575

5	77	31.6	428	9	AA199907	AA199907 zq52a07.r
6	76.4	31.3	604	10	BB655265	BB655265 BB655265
7	73	29.9	432	10	BB846631	BB846631 BB846631
8	63	25.8	1104	12	BM462913	BM462913 AGENCOURT
9	62	25.4	851	9	AU127876	AU127876 AU127876
10	55	22.5	773	9	AU132057	AU132057 AU132057
11	54	22.1	471	9	AA070261	AA070261 zm69c05.r
12	50	20.5	640	9	AU128573	AU128573 AU128573
13	44.6	18.3	661	14	CA415315	CA415315 UI-H-EZ0-
14	44	18.0	678	9	AL650594	AL650594 AL650594
15	43	17.6	550	9	AU129261	AU129261 AU129261
16	43	17.6	683	9	AU128985	AU128985 AU128985
17	41.4	17.0	913	13	EX408305	EX408305 EX408305
18	41.2	16.9	793	12	BG344149	BG344149 HVSMERG000
19	41	16.8	604	9	AU128022	AU128022 AU128022
20	40.6	16.6	1201	29	CNS01668	AL106346 Drosophil
21	39.2	16.1	465	14	CD333977	CD333977 StrPu536.
22	39	16.0	619	12	BI886285	BI886285 ZF637-1-0
23	38.2	15.7	899	13	EX349846	EX349846 EX349846
24	38	15.6	983	14	CA976262	CA976262 AGENCOURT
25	37.8	15.5	686	29	CNS0333T	AL225650 Tetraodon
26	37.8	15.5	1201	9	AL535176	AL535176 AL535176
27	37.4	15.3	853	29	AG138149	AG138149 Pan trogl
28	37.4	15.3	1033	28	AQ893007	AQ893007 HS_4832.A
29	37.4	15.3	1101	29	CNS016XU	AL107340 Drosophil
30	37.2	15.2	684	12	BI957311	BI957311 HVSMERG000
31	37	15.2	907	29	CG276221	CG276221 OG1DR10TV
32	37	15.2	1004	28	BZ559376	BZ559376 pacs2-164
33	36.8	15.1	819	29	CC907659	CC907659 t036e05ba
34	36.8	15.1	1021	28	CC225169	CC225169 CH261-114
35	36.6	15.0	469	28	BH254220	BH254220 SALK_0161
36	36.6	15.0	483	10	AW447419	AW447419 88968 MAR
37	36.6	15.0	602	29	CNS03KOU	AL247575 Tetraodon
38	36.6	15.0	603	12	BI074777	BI074777 IPL 15 B1
39	36.6	15.0	620	14	CB462398	CB462398 722478 MA
40	36.6	15.0	621	12	BI064787	BI064787 p9fln.pk0
41	36.6	15.0	1605	10	BF138648	BF138648 601781826
42	36.4	14.9	415	13	BY447104	BY447104 BY447104
43	36.4	14.9	469	12	BM508777	BM508777 i139908.Y
44	36.4	14.9	866	9	AL525171	AL525171 AL525171
45	36.4	14.9	893	10	BF122415	BF122415 601759708

## ALIGNMENTS

RESULT 1  
BE649209/c

LOCUS

DEFINITION

BE649209

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

BE649209 421 bp mRNA linear EST 06-SEP-2000  
UI-M-BH2.3-ach-f-07-0-UI.r1 NIH BMAP M S3.3 Mus musculus cDNA clone  
UI-M-BH2.3-ach-f-07-0-UI 5', mRNA sequence.

BE649209

BE649209.1 GI:9975033

EST.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 421)

Ronaldo,M.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

8889548

Contact: Chin, H

National Institute of Mental Health

6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD

20892-9643, USA

Tel: 301 443 1706

Fax: 301 443 9890

Email: mEST@mail.nih.gov

CDNA Library Preparation: M.B. Soares Lab Clone distribution:

Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements. The following repetitive elements were found in this cDNA sequence: 356-388, >GC\_rich#Low\_complexity

Seq primer: M13 Reverse

#### FEATURES

Location/Qualifiers

1..421  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UI-M-BH2.3-aoh-f-07-0-UI"  
 /dev\_stage="27-32 days"  
 /lab\_host="DH10B (Life Technologies)"  
 /clone\_lib="NIH BMAP M S3.3"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; The NIH\_BMAP\_M.S3.3 library is a subtracted library of a series, ultimately derived from a mixture of individually tagged normalized libraries from ten regions of the mouse brain (cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hippocampus) after a series of subtractions to reduce the representation of cDNAs from which ESTs had already been generated. The following serially subtracted libraries were generated in this process: NIH\_BMAP\_M.S3.3, NIH\_BMAP\_M.S2, NIH\_BMAP\_M.S1. The subtracted library (NIH\_BMAP\_M.S3.3) was constructed as follows: PCR amplified cDNA inserts from NIH\_BMAP\_M.S2 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the NIH\_BMAP\_M.S2 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the NIH\_BMAP\_M.S3.3 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)"

#### ORIGIN

Query Match 51.6%; Score 125.8; DB 10; Length 421;  
 Best Local Similarity 85.9%; Pred. No. 78-22; Indels 3; Gaps 1;  
 Matches 152; Conservative 0; Mismatches 22

QY 1 AGCGCGGGGATTTAAATGCCACTAAACGGGTGATCCATCTCGGAGCGCGCAACT 60  
 |||||  
 Db 343 AGCGCGGGGATTTAAATGCCACTGAACGGGTGATCCATCAGCGCGCGGTAACT 284  
 |||||

QY 61 TTTCGAGGAGGCTAGCCATTGGCTGACACCGTCACTGCGCCCTCTCCAGCGTCTCG 120  
 |||||

Db 283 TTTCGAGGCGGCTAGCCATTGGCTGGTAGTCAAGT---CCTCCCTCGAGCGCTCG 227  
 |||||

QY 121 CCTCCGCGCCCGCTCTGCGCACTGATCATCTATCATTTCTTCTCGGCCC 177  
 |||||

Db 226 CCTGCGCGCGCCGCTCTTGGCACTGATCATTCATATCTTCTCTGGTCC 170  
 |||||

#### RESULT 2

AW121149 343 bp mRNA linear EST 22-OCT-1999  
 LOCUS  
 DEFINITION UI-M-BH2.3-aoh-f-07-0-UI.s1 NIH\_BMAP\_M.S3.3 Mus musculus cDNA clone  
 UI-M-BH2.3-aoh-f-07-0-UI 3', mRNA sequence.  
 AW121149  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Mus musculus (house mouse)  
 Mus musculus  
 Fukuyota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 343)

#### AUTHORS

Bonaldo, M.F., Lennon, G. and Soares, M.B.  
 Normalization and subtraction: two approaches to facilitate gene discovery  
 Genome Res. 6 (9), 791-806 (1996)  
 97044477  
 889548  
 Contact: Chin, H  
 National Institute of Mental Health  
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD  
 20892-9643, USA  
 Tel: 301 443 1706  
 Fax: 301 443 9890  
 Email: mbest@mail.nih.gov  
 Oligo-dT track not found, Not I site shown in beginning of sequence is likely internal to the message. cDNA Library Preparation: M.B. Soares Lab Clone Distribution: NIH BMAP cDNA clones will be made available by the means that is soon to be determined. When NIH determines the means for distribution of the BMAP cDNA clones, this record will be updated accordingly when that means is determined. The following repetitive elements were found in this cDNA sequence: 13-50, >GC\_rich#Low\_complexity  
 Seq primer: M13 Forward  
 POLYA-No.

#### FEATURES

source

Location/Qualifiers

1..343  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UI-M-BH2.3-aoh-f-07-0-UI"  
 /dev\_stage="27-32 days"  
 /lab\_host="DH10B (Life Technologies)"  
 /clone\_lib="NIH\_BMAP\_M.S3.3"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; The NIH\_BMAP\_M.S3.3 library is a subtracted library of a series, ultimately derived from a mixture of individually tagged normalized libraries from ten regions of the mouse brain (cerebellum, brain stems, olfactory bulbs, pineal hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hippocampus) after a series of subtractions to reduce the representation of cDNAs from which ESTs had already been generated. The following serially subtracted libraries were generated in this process: NIH\_BMAP\_M.S3.3, NIH\_BMAP\_M.S2, NIH\_BMAP\_M.S1. The subtracted library (NIH\_BMAP\_M.S3.3) was constructed as follows: PCR amplified cDNA inserts from NIH\_BMAP\_M.S2 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the NIH\_BMAP\_M.S2 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the NIH\_BMAP\_M.S3.3 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)  
 TAG\_TISSUE=brain-stems  
 TAG\_LIB=NIH\_BMAP\_M.S3.3  
 TAG\_SEQ=TCATG"

#### ORIGIN

Query Match 50.9%; Score 124.2; DB 10; Length 343;  
 Best Local Similarity 85.3%; Pred. No. 1.7e-21; Indels 3; Gaps 1;  
 Matches 151; Conservative 0; Mismatches 23

QY 1 AGCGCGGGGATTTAAATGCCACTAAACGGGTGATCCATCTCGGAGCGCGCAACT 60  
 |||||

Db 66 AGCGCGGGGATTTAAATGCCACTGAACGGGTGATCCATCAGCGCGCGGTAACT 125  
 |||||

QY 61 TTTCGAGGAGGCTAGCCATTGGCTGACACCGTCACTGCGCCCTCTCCAGGTCCTCG 120  
 |||||

Db 126 TTTCGATGCGGCTAGCCATTGGCTGGTAGTCAAGT---CCTCCCTCGAGCGCTCG 182  
 |||||

```

QY 121 CCTCTCCCGCCCGCCCTCTTGCGCACTGTACATTCATATCAATTTCTTCGCGCC 177
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 183 CCTGCGCGCCCGCCCTCTTGCGCACTGTACATTCATATCAATTTCTTCGCGCC 239

RESULT 3
LOCUS BM838150
DEFINITION K-EST0114344 S9SNU601 Homo sapiens cDNA clone S9SNU601-73-G08 5',
            mRNA sequence.
ACCESSION BM838150
VERSION BM838150.1 GI:19194559
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 625)
AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
TITLE 21c Frontier Korean EST Project 2001
JOURNAL Unpublished (2002)
COMMENT Contact: Kim YS
        Genome Research Center
        Korea Research Institute of Bioscience & Biotechnology
        52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
        Tel: +82-42-860-4470
        Fax: +82-42-860-4409
        Email: yongsung@mail.kribb.re.kr
        Plate: 73 row: G column: 08
        High quality sequence stop: 625.
        Location/Qualifiers
            1..625
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="S9SNU601-73-G08"
                /sex="M"
                /tissue_type="Ascites"
                /cell_type="Epithelial"
                /cell_line="SNU-601"
                /lab_host="Top10F"
                /clone_lib="S9SNU601"
                /note="Organ: Stomach; Vector: pME18-FL3; Site 1: XhoI;
                Site 2: XhoI; The poly (A)+ RNA was dephosphorylated with
                bacterial alkaline phosphatase (BAP) and then decapped
                with tobacco acid pyrophosphatase (TAP). The decapped
                intact mRNA was ligated with DNA-RNA linker including SfiI
                site by treatment of T4 RNA ligase and the first strand
                cDNA was synthesized with Superscript II using SfiI
                oligo-dr primer. After first strand synthesis, RNA was
                degraded by NaOH treatment and cDNA was amplified by PCR
                reaction. The PCR products were digested with SfiI and
                cloned into DraIII- digested pME18S-FL3 vector. The
                obtained cDNA vectors were used for transformation of
                competent cells E. coli Top10F by electroporation method.
                The cDNA libraries constructed by this method are
                full-length enriched cDNA library."

ORIGIN
Query Match 36.1%; Score 88; DB 12; Length 625;
Best Local Similarity 100.0%; Pred. No. 4.1e-12;
Matches 88; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 157 ATCATTTTCTCTCCGCGCCCGCCATGGAGGAAGTGAGAAAGTTGCGACAGTCACGCGCGC 216
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 ATCATTTTCTCTCCGCGCCCGCCATGGAGGAAGTGAGAAAGTTGCGACAGTCACGCGCGC 60

QY 217 TTGCGAGGACCAAGTCACTCAGTGACAG 244
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 TTGCGAGGACCAAGTCACTCAGTGACAG 88

```

```

RESULT 4
LOCUS AK083575
DEFINITION Mus musculus 9 days embryo whole body cDNA, RIKEN full-length
            enriched library, clone:D030047C07 product:homeo box A1, full
            insert sequence.
ACCESSION AK083575
VERSION AK083575.1 GI:26350638
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 93279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6
AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,M.,
Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
Okazaki,Y., Saito,K., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akashira,S.,
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222.

```

COMMENT  
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
 Please visit our web site for further details.  
 URL: <http://genome.gsc.riken.go.jp/>  
 URL: <http://fantom.gsc.riken.go.jp/>

FEATURES  
 source  
 1..2475  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="FANTOM DB:D030047C07"  
 /db\_xref="MGI:2418936"  
 /db\_xref="taxon:10090"  
 /clone="D030047C07"  
 /tissue\_type="whole body"  
 /clone\_lib="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="9 days embryo"  
 98..1108  
 /note="unnamed protein product; homeo box A1 (MGD|MGI:96170, GB|NM\_010449, evidence: BLASTN, 99%, match=2216)  
 putative"  
 /codon\_start=1  
 /protein\_id="BAC38956.1"  
 /db\_xref="GI:26350639"  
 /translation="MDNARMSFLEYPILGSDGTCGARAYPSDHGITTFOCAVSA  
 RSCGGDPLVGRVQISPHHHHHHHHPOTATYOTSGNLGISYSHSCGPGYGA  
 QNFSAPYGGYGLNQADVGSGPPCAPAVYSGNLSTPMVQHHRHOGVAGTGVGSPQY  
 IHHSYQBOQTALATYNNLSPLHSHOEACRSPASSTSPQAFTDMKVKRNPPKT  
 KWFGYGVYQGNVATNFNTKLTLEKEHFNKYLTRARVEIAASLQNLNETQVKI  
 WFNRRMKQKKEKGLLPISPATPPGSDKTESEKSPSPSPASSTSDTLTT  
 SH"

CDS  
 32.8%; Score 80; DB 11; Length 2475;  
 Best Local Similarity 89.6%; Pred. No. 6.6e-10;  
 Matches 86; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

ORIGIN  
 149 ACATTCATATCATTTTCTCTCCGCCCCCGAGGAGTGGAGAAAGTGGACAGTCA 208  
 Db 2 ACATTCATATCATTTTCTCTCTCTGTCCTATGAGGAAGTGGACAGTCA 61  
 209 CGCGGGCTTCGAGGACGAGTCACTCAGTGACAG 244  
 Db 62 CCAGGGCTTCGAGGATCCATCATCTCAGTGACAG 97

RESULT 5  
 LOCUS  
 DEFINITION  
 AAL199907 428 bp mRNA linear EST 23-JAN-1997  
 z952a07.r1 Stratagene neuroepithelium (#937231) Homo sapiens cDNA  
 clone IMAGE:645204 5', similar to TF:G1195540 G1195540 HOXA1  
 {ALTERNATIVELY SPLICED} . ;, mRNA sequence.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 AAL199907.1 GI:1795641  
 Homo sapiens (human)

REFERENCE  
 AUTHORS  
 Hillier, L., Lennon, G., Becker, M., Bonaldo, M.P., Chiapelli, B.,  
 Chissoe, S., Dietrich, N., Duboue, T., Favello, A., Gish, W.,  
 Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,  
 Marra, M., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,  
 Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierri-Meg, J.,  
 Trevaskis, E., Underwood, K., Wohldmann, P., Waterston, R., Wilson, R.,  
 and Marra, M.

TITLE  
 Generation and analysis of 280,000 human expressed sequence tags

JOURNAL  
 MEDLINE  
 PUBMED  
 COMMENT  
 Genome Res. 6 (9), 807-828 (1996)  
 97044478  
 889549  
 Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: [est@watson.wustl.edu](mailto:est@watson.wustl.edu)  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
 Seq primer: -28M13 rev2 from Amersham  
 High quality sequence stop: 195.

FEATURES  
 source  
 1..428  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="GDB:5215455"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:645204"  
 /dev\_stage="Ntera-2/RA neuroepithelial cells"  
 /lab\_host="SOLR (kanamycin resistant)"  
 /clone\_lib="Stratagene neuroepithelium (#937231)"  
 /note="Vector: pBluescript SK-; Site: 1: EcoRI; Site 2:  
 XhoI; Cloned unidirectionally. Primer: Oligo dt. NT2  
 cells (Ntera-2/cl.D1) induced with Retinoic Acid for 24  
 hours. Average insert size: 1.5 kb; Uni-ZAP XR Vector; ~5'  
 adaptor sequence: 5' GAATTCGACGACG 3' ~3' adaptor  
 sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"

ORIGIN  
 Query Match 31.6%; Score 77; DB 9; Length 428;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-09;  
 Matches 77; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

168 TCTCGGGCCCCATGAGGAAAGTGGACAGTTCGACAGTCACTCAGTGACGAC 227  
 QY |||||  
 Db 1 TCTCGGGCCCCATGAGGAAAGTGGACAGTTCGACAGTCACTCAGTGACGAC 60  
 228 AGTCACTCAGTGACAG 244  
 QY |||||  
 Db 61 AGTCACTCAGTGACAG 77

RESULT 6  
 LOCUS  
 DEFINITION  
 BB655265 604 bp mRNA linear EST 26-OCT-2001  
 BB655265 RIKEN full-length enriched, 9 days embryo Mus musculus  
 cDNA clone D030047C07 5', mRNA sequence.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 BB655265.1 GI:16489093  
 Mus musculus (house mouse)

REFERENCE  
 AUTHORS  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 604)  
 Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T.,  
 Hara, A., Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J.,  
 Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K.,  
 Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K.,  
 Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
 Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,  
 Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M. and Hayashizaki, Y.  
 RIKEN Mouse ESTs (Arakawa, T., et al. 2001)

TITLE  
 JOURNAL  
 COMMENT  
 Unpublished (2001)  
 Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic  
 Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216

## ORIGIN

```

Query Match      29.9%; Score 73; DB 10; Length 432;
Best Local Similarity 88.8%; Pred. No. 2.7e-08;
Matches 79; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 156 TATCATTTTCTTCCTCCGGCCCATGGAGGAAGTGAGAAAGTTGGCAGAGTCAGCCGGG 215
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3 TATCATTTTCTTCCTCTGCTCTATGAGGAAGTGAGAAAGTTGGCAGCGTCACCCAGG 62
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 216 CTTCCAGGACCAAGTCACTCAGTGACAG 244
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 63 CTTCCAGGATCCATCACTCAGTGACAG 91
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 8
BM462913
LOCUS
DEFINITION AGENCOURT_6447587 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5537563
5', mRNA sequence.
ACCESSION BM462913
VERSION
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC/DCTD/DTP
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L7AM1229 row: a column: 20
High quality sequence stop: 717.
Location/Qualifiers
1..1104
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5537563"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 72"
/Note="Organ: Skin; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2 kb. Library constructed by Life
Technologies."

ORIGIN
Query Match      25.8%; Score 63; DB 12; Length 1104;
Best Local Similarity 100.0%; Pred. No. 1.3e-05;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 182 GAGGAAGTGAGAAAGTTGGCAGAGTCAGCCGGGCTTCGAGGACCAAGTCACTCAGTGA 241
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 11 GAGGAAGTGAGAAAGTTGGCAGAGTCAGCCGGGCTTCGAGGACCAAGTCACTCAGTGA 70
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 242 CAG 244
      |||||
Db 71 CAG 73

RESULT 9
AU127876
LOCUS
DEFINITION AU127876 NT2RP2 Homo sapiens cDNA clone NT2RP2002264 5', mRNA
sequence.
ACCESSION AU127876
VERSION
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE HRI human cDNA project
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan

```

```

sequence.
ACCESSION AU127876
VERSION AU127876.1 GI:10988230
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 851)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
TITLE HRI human cDNA project
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
Location/Qualifiers
1..851
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="NT2RP2002264"
/cell_type="teratocarcinoma"
/clone_lib="NT2"
/clone_libs="NT2RP2"
/Note="Vector: pME18SFL3; mRNA from NT2 neuronal precursor
cells after 2-weeks retinoic acid (RA) induction"

ORIGIN
Query Match      25.4%; Score 62; DB 9; Length 851;
Best Local Similarity 100.0%; Pred. No. 2.1e-05;
Matches 62; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 183 AGGAAGTGAGAAAGTTGGCAGAGTCAGCCGGGCTTCGAGGACCAAGTCACTCAGTGAC 242
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 AGGAAGTGAGAAAGTTGGCAGAGTCAGCCGGGCTTCGAGGACCAAGTCACTCAGTGAC 60
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 243 AG 244
      |||||
Db 61 AG 62

RESULT 10
AU132057
LOCUS
DEFINITION AU132057 NT2RP3 Homo sapiens cDNA clone NT2RP3003711 5', mRNA
sequence.
ACCESSION AU132057
VERSION AU132057.1 GI:10992411
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 773)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
TITLE HRI human cDNA project
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan

```



Tel: 81-438-52-3975  
Fax: 81-438-52-3986

Email: genomics@hri.co.jp  
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.

## FEATURES

source

1. .773  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="NT2RP3003711"  
/cell\_type="teratocarcinoma"  
/clone\_lib="NT2"  
/clone\_lib="NT2RP3"  
/notes="Vector: pME18SFL3; mRNA from NT2 neuronal precursor cells after 2-weeks retinoic acid (RA) induction"

## ORIGIN

Query Match 22.5%; Score 55; DB 9; Length 773;  
Best Local Similarity 100.0%; Pred. No. 0.0013;  
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 190 GAGAAAGTTGGCAGTCACGCCGGGCTTCGCAGGACCAGGTCACTCAGTGACAG 244  
|||||  
Db 1 GAGAAAGTTGGCAGTCACGCCGGGCTTCGCAGGACCAGGTCACTCAGTGACAG 55

## RESULT 11

AA070261

LOCUS

DEFINITION

zm69c05.r1 Stratagene neuroepithelium (#937231) Homo sapiens cDNA clone IMAGE:530888 5' similar to SW:HXAI\_MOUSE P09022 HOMEOBOX PROTEIN HOX-A1 ; mRNA sequence.

ACCESSION

AA070261

VERSION

EST.

SOURCE

ORGANISM

REFERENCE

AUTHORS

Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1. (bases 1 to 471)  
Hallier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chisapelli, B., Chisoe, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, P., Thierry-Mieg, J., Trevas, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R., and Warra, M.  
Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

Contact: Wilson RK  
Washington University School of Medicine  
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

WARNING: There is evidence that suggests that the 384-well parent plate of this clone contains both human and mouse derived clones. Thus, the origin of this clone is uncertain. This caution should be kept in mind should you use this clone.

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.

Seq primer: -28M13 rev2 from Amer-sham

High quality sequence stop: 440.

## FEATURES

source

1. 471  
/organism="Homo sapiens"  
/mol\_type="mRNA"

/db\_xref="GDB:3920400"  
/db\_xref="taxon:9606"  
/clone="IMAGE:530888"  
/dev\_stage="Ntera-2/RA neuroepithelial cells"  
/lab\_host="SOLR (kanamycin resistant)"  
/clone\_lib="Stratagene neuroepithelium (#937231)"  
/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer: Oligo dt. NT2 cells (Ntera-2/cl.D1) induced with Retinoic Acid for 24 hours. Average insert size: 1.5 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGCACGAG 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTT 3'"

## ORIGIN

Query Match 22.1%; Score 54; DB 9; Length 471;  
Best Local Similarity 100.0%; Pred. No. 0.0021;  
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 191 AGAAAGTTGGCAGTCACGCCGGGCTTCGCAGGACCAGGTCACTCAGTGACAG 244  
|||||  
Db 1 AGAAAGTTGGCAGTCACGCCGGGCTTCGCAGGACCAGGTCACTCAGTGACAG 54

## RESULT 12

AA128573

LOCUS

DEFINITION

640 bp mRNA linear EST 01-AUG-2002  
AA128573 NT2RP2 Homo sapiens cDNA clone NT2RP2003721 5', mRNA sequence.

ACCESSION

AA128573

VERSION

EST.

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

RESULT 13  
CA415315  
LOCUS CA415315 661 bp mRNA linear EST 07-NOV-2002  
DEFINITION UI-H-EZO-baw-j-13-0-UI.s1 NCI\_CGAP Ch1 Homo sapiens cDNA clone  
UI-H-EZO-baw-j-13-0-UI 3', mRNA sequence.  
ACCESSION CA415315  
VERSION CA415315  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 661)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
CONTACT: Robert Strausberg, Ph.D.  
Email: [cgaps@mail.nih.gov](mailto:cgaps@mail.nih.gov)  
Tissue Procurement: Dr. Steven Gitelis/ Rush Presbyterian, Dept. of  
Orthopaedics  
CNA Library preparation: Dr. M. Bento Soares, University of Iowa  
CNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Clone distribution information can be obtained  
from Dr. M. Bento Soares, [bento-soares@uiowa.edu](mailto:bento-soares@uiowa.edu)  
The following repetitive elements were found in this cDNA  
sequence: 1-43, >POLY(A)simple\_repeat (matched complement) 413-491,  
>GC-rich#low complexity  
Seq primer: M13 FORWARD  
POLYA=Yes.

FEATURES  
source  
1..661  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-H-EZO-baw-j-13-0-UI"  
/tissue\_type="Chondrosarcoma Grade II"  
/dev\_stage="Adult"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="NCI CGAP Ch1"  
/note="Organ: Left Pelvis; Vector: pVT73-Pac (Pharmacia)  
with a modified polylinker; Site 1: EcoR I; Site 2: Not I;  
NCI CGAP Ch1 is a cDNA library containing the following  
tissue(s): Chondrosarcoma Grade II. The library was  
constructed according to Bonaldo, Lennon and Soares,  
Genome Research, 6:791-806, 1996. First strand cDNA  
synthesis was primed with an oligo-dT primer containing a  
Not I site. Double stranded cDNA was ligated to an EcoR I  
adaptor, digested with Not I, and cloned directionally  
into pVT73-Pac vector. The oligonucleotide used to prime  
the synthesis of first-strand cDNA contains a library tag  
sequence that is located between the Not I site and the  
(dT)18 tail. The sequence tag for this library is  
TGATCAGCGT.  
TAG\_TISSUE=grade-2-chondrosarcoma  
TAG\_LIB=UI-H-EZO  
TAG\_SEQ=ATCTATATG

ORIGIN  
Query Match 18.3%; Score 44.6; DB 14; Length 661;  
Best Local Similarity 92.2%; Pred. No. 0.6;  
Matches 47; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
Qy 1 AGCCCGGGGATTTAAATGCCACTAAACGGTGTCCATCTCTCGGAAGC 51  
Db 610 AGCCCGGGGATTTAAATGCCACTAAACGGTGTCCATCTCTCGGAGC 660  
RESULT 14  
AL650594  
LOCUS AL650594 678 bp mRNA linear EST 07-NOV-2003

DEFINITION AL650594 XGC-neurula Silurana tropicalis cDNA clone TNeu028f22 5',  
mRNA sequence.  
ACCESSION AL650594  
VERSION AL650594.2 GI:38223211  
KEYWORDS EST.  
SOURCE Silurana tropicalis (western clawed frog)  
ORGANISM Silurana tropicalis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;  
Xenopodinae; Silurana.  
REFERENCE 1 (bases 1 to 678)  
AUTHORS Croning, M.D.R., Ashurst, J.L., Taylor, R., Zorn, A.M. and Rogers, J.  
TITLE Sanger Xenopus tropicalis EST project 2001 (11\_2003)  
JOURNAL Unpublished (2003)  
COMMENT On Dec 13, 2001 this sequence version replaced gi:17660265.  
Contact: Huckle E  
Sanger Institute  
Hinxton, Cambridgeshire, CB10 1SA, UK  
Email: [trop@sanger.ac.uk](mailto:trop@sanger.ac.uk)  
This sequence is from a Xenopus Gene Collection (XGC) library  
constructed by Aaron M. Zorn.  
cDNA was oligo dt primed from Sug of poly A+ RNA from neurula.  
EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the  
5' end and NotI at the 3' end.  
Vector: pCS107; Site 1: EcoRI; Site 2: NotI  
Host: Escherichia coli DH10B  
Sanger Xenopus tropicalis EST project 2001  
TROPICALIS SEQUENCE ID: TNeu028f22.plkSP6  
Sequencing primer: SP6.  
FEATURES  
Location/Qualifiers  
1..678  
/organism="Silurana tropicalis"  
/mol\_type="mRNA"  
/db\_xref="taxon:8364"  
/clone="TNeu028f22"  
/dev\_stage="neurula"  
/lab\_host="Escherichia coli DH10B"  
/clone\_lib="XGC-neurula"  
/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA  
was oligo dt primed from Sug of poly A+ RNA from neurula.  
EcoRI-NotI cut cDNA was then ligated into pCS107 with  
EcoRI at the 5' end and NotI at the 3' end."

ORIGIN  
Query Match 18.0%; Score 44; DB 9; Length 678;  
Best Local Similarity 72.0%; Pred. No. 0.86;  
Matches 72; Conservative 0; Mismatches 25; Indels 3; Gaps 1;  
Qy 118 CGGCCCTCCGCCGCCCTCTTGGCACTGTACATTCATATCATTTTCTTCGCGCC 177  
Db 103 CTGGCCTCTCCACGTGACCTGCCCTCCTACTGTACATTCATATGATT---TGTGCTGGACC 159  
Qy 178 CATGGAGGAAGTGGAGAAAGTTGGCACAGTCACGCCGGCT 217  
Db 160 TATGGAGGAAGTGGAGAAAGTTGGCACCGCCACGCCAGGCT 199

RESULT 15  
AUI29261  
LOCUS AUI29261 550 bp mRNA linear EST 01-AUG-2002  
DEFINITION AUI29261 NT2RP2 Homo sapiens cDNA clone NT2RP2005190 5', mRNA  
sequence.  
ACCESSION AUI29261  
VERSION AUI29261.1 GI:10989615  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 550)  
AUTHORS Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y.,  
Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagai, T., Sugano, S. and  
Isogai, T.

TITLE HRI human cDNA project  
JOURNAL Unpublished (2000)  
COMMENT Contact: Takao Isogai  
Genomics Laboratory  
Helix Research Institute  
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3975  
Fax: 81-438-52-3986  
Email: genomics@hri.co.jp

HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix  
Research Institute; cDNA library construction: Department of  
Virology, Institute of Medical Science, University of Tokyo, and  
Helix Research Institute.

# FEATURES

source

1. .550  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="NT2RP2005190"  
/cell\_type="teratocarcinoma"  
/cell\_line="NT2"  
/clone\_lib="NT2RP2"  
/note="Vector: pME18SFL3; mRNA from NT2 neuronal precursor  
cells after 2-weeks retinoic acid (RA) induction"

# ORIGIN

Query Match 17.6%; Score 43; DB 9; Length 550;  
Best Local Similarity 100.0%; Pred. No. 1.5;  
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 202 ACAGTCACGCCGGCTTCGAGGACCAGGTCAGTCAGTGACAG 244  
|||||  
Db 1 ACAGTCACGCCGGCTTCGAGGACCAGGTCAGTCAGTGACAG 43  
|||||

Search completed: September 18, 2004, 13:16:52  
Job time : 2065.12 secs

**This Page Blank (uspto)**